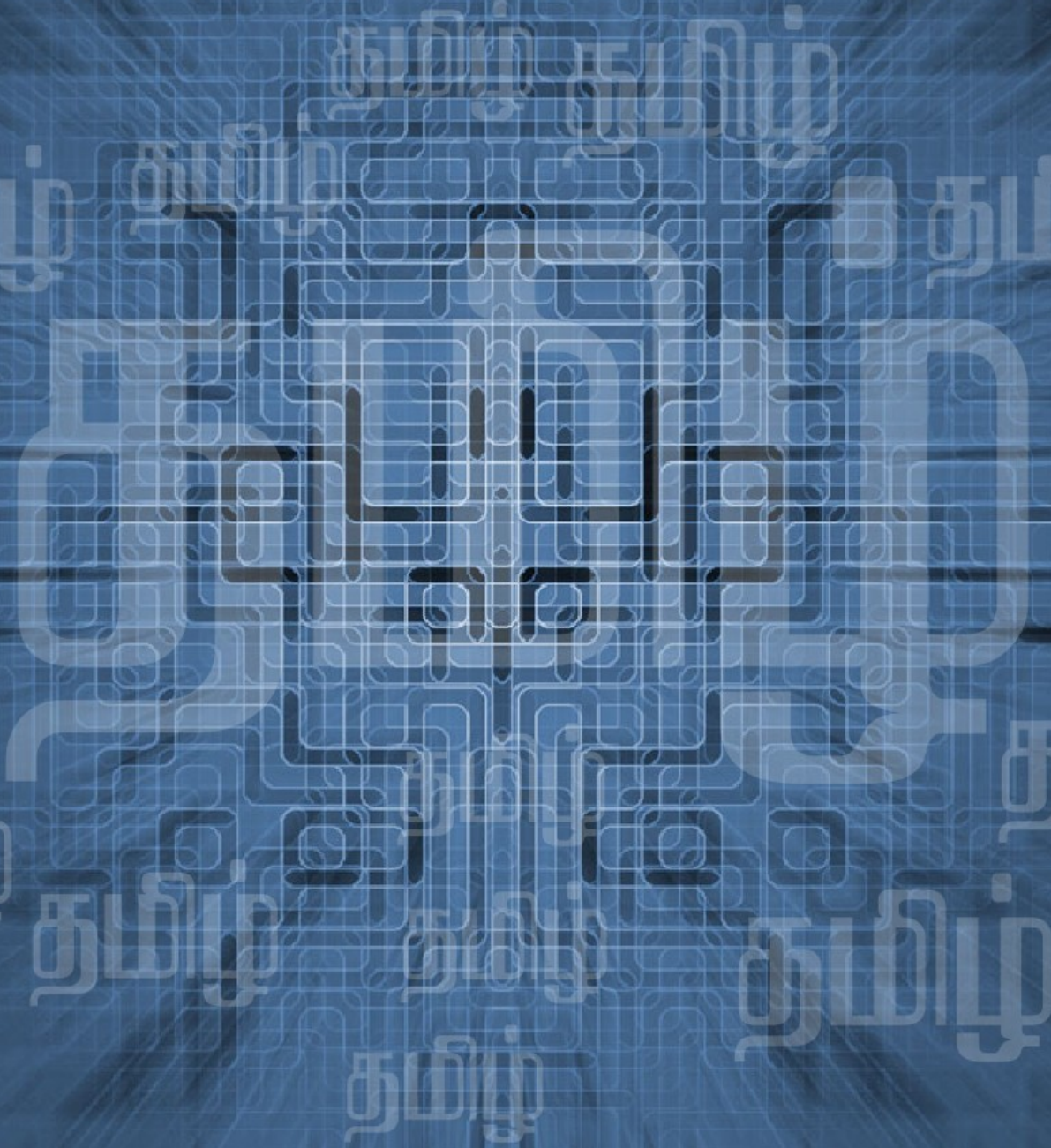


தமிழின் எதிர்காலமும் தகவல் தொழில்நுட்பமும்



இரா.அசோகன்

தமிழின் எதிர்காலமும் தகவல் தொழில்நுட்பமும்

இரா. அசோகன்

ashokramach@gmail.com

அட்டைப்படம் : லெனின் குருசாமி - guruleninn@gmail.com

மின்னுலாக்கம் : சி.ராஜேஸ்வரி - sraji.me@gmail.com

வெளியிடு : FreeTamilEbooks.com

உரிமை – Creative Commons Attribution-ShareAlike CC BY-SA

பொருளடக்கம்

1. தமிழின் எதிர்காலமும் தகவல் தொழில்நுட்பமும் — 1.....	4
2. தொடர்ந்து இரண்டு ஆங்கில மொழிப் பேரரசுகள்.....	7
3. உலகமயமாக்கலும் தகவல் தொழில்நுட்பமும்.....	10
4. அமெரிக்க கலாச்சார ஆதிக்கமும் தகவல் தொழில்நுட்பமும்.....	12
5. பரவும் இந்த ஆங்கில மொழி ஆதிக்கம் தணிய வாய்ப்பு உள்ளதா?.....	15
6. தடை செய்யப்பட்ட கட்டலான் மொழி புத்துயிர் பெற்றது எப்படி?.....	18
7. “ஆளும் மொழியே வாழும்; மற்றவை மாளும்” இதுதான் நியதியா?.....	21
8. புதிய தலைமுறையின் மரபணுவே எண்ணிமத்தால் ஆனது போலுள்ளது.....	24
9. உங்கள் பிள்ளைகளை இயந்திர மனிதர்களாக வளர்க்கிறீர்களா?.....	27
10. கணினிக்குத் தமிழ் கற்றுக் கொடுப்போம் வாருங்கள்.....	30
11. பெரு நிறுவனங்களின் தமிழ் சேவைகளை நம்பியே இருந்தால் என்ன?.....	33
12. ஏன் திறந்த மூலமும், திறந்த தரவுகளும், திறந்த ஆய்வும்?.....	36
13. நிரல் எழுதத் தெரியாதவர்கள் 21 ஆம் நூற்றாண்டின் தற்குறிகளா?.....	39
14. உங்கள் பிள்ளைகளின் கணினி ஐபேடா, ராஸ்ப்பெரி-பையா?.....	42
15. தகவல் தொழில்நுட்ப உதவியுடன் தமிழை எளிதாகக் கற்பிக்கலாம்.....	45
16. தகவல் தொழில்நுட்ப உதவியுடன் நீங்களும் ஒரு எழுத்தாளராகலாம்.....	48
17. உரையும் பேச்சும் கொண்ட மொழித்தொகுப்பு.....	51
18. சொல்வகைக் குறியீடும் குறியிட்ட உரைத்தொகுப்புகளும்.....	54
19. வாக்கியக் கூறு பிரித்தலும், பெயரிட்ட உருபொருள் அடையாளம் காணுதலும்.....	57
20. தமிழின் தனித்தன்மைகளை வைத்துக் குறியிட்ட உரைகள் தேவையைக் குறைக்க முடியுமா?....	60
21. சொல்வையும் சொல்லின் பொருளில் ஐயமகற்றலும்.....	63
22. அடிச்சொல், தண்டுச்சொல் மற்றும் சொற்பகுப்பாய்வு.....	66
23. சார்புநிலை பிரிப்பி, சுட்டுப்பெயர் தீர்வு, தலைப்பு பிரித்தெடுத்தல்.....	69
24. இயல்மொழி ஆய்வு கருவித் தொடரி.....	72
25. தமிழ் — ஆங்கிலம் இயந்திர மொழிபெயர்ப்பு.....	76
26. சொற்பிழைத் திருத்தி.....	79
27. உணர்வு பகுப்பாய்வும் சமூக ஊடகங்களும்.....	82
28. மொழித் தொழில்நுட்பத்தில் வளங்கள் மிகுந்த மொழியாகத் தமிழை உயர்த்துவோம்.....	85

1. தமிழின் எதிர்காலமும் தகவல் தொழில்நுட்பமும்

- 1

"தமிழின் நிகழ்காலம் குறித்து வருந்துகிறேன். எதிர்காலம் குறித்து அஞ்சுகிறேன்"

தமிழின் நிகழ்காலப் போக்கு

தமிழ் இன்று அறிவியல் மற்றும் ஆராய்ச்சி மொழியாகவும், உயர் கல்வி மற்றும் வணிக மொழியாகவும் இல்லை. ஆகவே உயர் கல்வி மற்றும் ஆராய்ச்சியிலும், பணியிடத்திலும் ஓரளவாவது ஆங்கிலத்தில் பரிச்சயம் இல்லையெனில் யாரும் சமாளிக்க முடியாது என்பது மறுக்க முடியாத உண்மைதான். மேலும் விற்பனை மற்றும் சேவைத் துறையில் பணிபுரிந்தாலோ, மென்பொருள் உருவாக்குவோரும் மற்றவர்களும் வெளிநாட்டவருடன் சேர்ந்து வேலைசெய்ய வேண்டியிருந்தாலோ அல்லது தகவல் அழைப்பு மையத்தில் வெளிநாட்டவர் தொலைபேசி அழைப்புகளுக்குப் பதில் சொல்ல வேண்டுமென்றாலோ சரளமாக ஆங்கிலம் பேச எழுத முடிந்தாலொழிய வேலை செய்ய முடியாது.

ஆனால் ஆங்கிலத் தொடர்புள்ள எதையும் மேம்பட்டதென்றும், சிறப்பானதென்றும், நவநாகரிகமானதென்றும் பலரும் நினைக்கிறார்கள். மழலையர் பள்ளியில் சேர குழந்தையின் ஆங்கிலத் திறனை மேம்படுத்துவதற்காகப் பெற்றோர்கள் தங்கள் குழந்தைகளை வீட்டில் ஆங்கிலத்தில் மட்டுமே பேசவைக்குமாறு பல பள்ளிகள் வற்புறுத்துகின்றன. பிற்காலத்தில் வேலைவாய்ப்பில் பின்தங்கிவிடாமல் இருக்க, பெற்றோர்களும் இந்த அறிவுரைகளைத் தீவிரமாகப் பின்பற்றுகிறார்கள். இதன் விளைவாகக் குழந்தை ஆங்கிலம் எழுதவும், பேசவும் நன்றாகக் கற்றுக் கொள்ளலாம். ஆனால் தன் தாய்மொழியில் ஆர்வமற்றுப் போகிறது. மொழி என்பது தொடர்பு கொள்வதற்கான வழி மட்டுமல்ல. கலாச்சார, வட்டார மற்றும் தேசிய அடையாளத்துக்கு மொழி மிகவும் வலுவான நங்கூரமாகவும் உள்ளது. மொழி என்பது அம்மொழி பேசும் மக்களின் மனப்பாங்கு மற்றும் அவர்கள் கூட்டாக எதை மதிக்கிறார்கள் என்பதன் பிரதிபலிப்பும் தான்.

இணையத்தில் ஒருவர் அங்கலாய்க்கிறார், “இது சோகமானது ஆனால் உண்மை. ஓர் இந்தியனாக நான் பார்க்கிறேன் உண்மையில் ஒரு நபரின் திறமையை எப்படிப் பலர் அளவிடுகிறார்கள் என்று – ஆங்கில மொழியில் அவர்களால் சரளமாகப் பேச முடிகிறதா என்பதை வைத்து.” மற்றொருவர் சொல்கிறார், “நான் ஆங்கில மொழி பிடித்தவர்களை எதிர்ப்பவன் அல்ல. ஆனால் ஆங்கிலம் அல்லாத மொழி பேசுபவர்கள் எல்லோருக்கும் குறைவான எழுத்தறிவு இருப்பதாக எண்ணுவதை எதிர்க்கிறேன்.” ஆனால் இம்மாதிரி மனப்பாங்கு உள்ளவர்கள் குறைந்து வருகிறார்கள் என்பதும் குறிப்பாக இளைய தலைமுறையில் கவலைதரும் அளவில் குறைந்து வருகிறார்கள் என்பதும் வருந்தத் தக்க உண்மைதான்.

தமிழின் எதிர்காலத்தை ஏன் இருள் சூழ்ந்து விட்டது?

மேற்கண்டவாறு முந்தைய தலைமுறைகளுடன் ஒப்பிடும்போது இன்றைய தலைமுறையினரின் உளப்பாங்குகளும் அணுகுமுறையும் மாறியுள்ளதும் கவனிக்க வேண்டியுள்ளது. இந்தத் தலைமுறையின் உளப்பாங்கும் அணுகுமுறையும் ஆழமாகத் தெரிய தமிழ் இளைஞர் மேற்கோள் எதுவும் நேரடியாகக் கிடைக்கவில்லை. ஆனால் இந்தச் சீன இளைஞரின் மேற்கோள் இந்தச் சந்தர்ப்பத்தில் பொருத்தமானதே.

“நான் ஒரு மொழியை நாணயம்போலப் பார்க்கிறேன். மொழிகளின் உலகில் ஆங்கில மொழி நாணயங்களின் உலகில் அமெரிக்க டாலர்கள் போன்றது. அதிகம் பேர் அதனைப் பயன்படுத்துவதால் அதன் மதிப்பு அதிகரிக்கிறது. ஆங்கிலம் ஏற்கனவே ஒரு பெரிய பிணைய விளைவுகளைக் (network effect) கொண்டுள்ளது. உலகமயமாக்கல் மேலும் ஆகிக்கொண்டிருப்பதால் இந்த நன்மையும் மேலும் அதிகரிக்கத்தான் செய்யும். எனக்கு ஆங்கிலம் சரளமாகத் தெரியவரும் முன்னரே சீன மொழியில் எழுதப்பட்ட புத்தகங்களை, என் தாய் மொழியாக இருந்த போதிலும், படிப்பதை நிறுத்தி விட்டேன். காரணம் ஆங்கிலத்துடன் ஒப்பிடுகையில் சீன மொழியில் எழுதப்பட்ட நூல்களில் கருத்துக்களின் வறுமைதான்.” இம்மாதிரி இளைய தலைமுறையினர் தங்களுடைய வேலை வாய்ப்புக்கும் தொழில்

வாழ்க்கைக்கும் எது முக்கியம் என்று பார்த்து அதில் மட்டும் கவனம் செலுத்துகிறார்கள். வழியில் தாய்மொழியும், பண்பாடும், கலாச்சாரமும் உடன் சேதம் (collateral damage) ஆகி விடுகின்றன. இணைய தலைமுறையைக் குறை சொல்லும் அதே நேரத்தில், அவர்களுடைய பெற்றோர்களாகிய நாமும்தான் பல சந்தர்ப்பங்களில் இதற்கு உடந்தையாகவும் தூண்டுபவர்களாகவும் இருக்கிறோம், கல்வி மற்றும் வேலையில் கவனம் முக்கியமென்றும் வேறெதிலும் திசை திருப்ப வேண்டாம் என்றும் வற்புறுத்துவதனால், என்பதையும் மறந்து விடக் கூடாது.

“தமிழின் நிகழ்காலம் குறித்து வருந்துகிறேன். எதிர்காலம் குறித்து அஞ்சுகிறேன்.”

2015 இல் ஒரு நேர்காணலில் கவிஞர் வைரமுத்துவைக் கேட்கிறார்கள், “இன்றைய தலைமுறையினரிடையே தமிழ்ப் பற்றும், இலக்கிய நாட்டமும் உள்ளதா?” என்று. அவர் கூறிய பதில் இது, “தமிழின் நிகழ்காலம் குறித்து வருந்துகிறேன். எதிர்காலம் குறித்து அஞ்சுகிறேன். இது தமிழுக்கு மட்டுமான பின்னடைவாக இல்லாமல் தேசிய மொழிகள் அனைத்துக்குமான பின்னடைவாகக் கருதுகிறேன். ஒரு தமிழன் என்பதால் தமிழ் மொழி குறித்துக் கூடுதல் கவலை அடைகிறேன். எழுத்து, பேச்சு என்ற இரண்டு வடிவங்களில்தான் ஒரு மொழி நிலை கொள்கிறது. தமிழ் படித்தால் என்ன பயன் என்று கருதுகிற ஒரு கூட்டம் தங்கள் பிள்ளைகளை ஆங்கில வழிக் கல்விக்கு ஆற்றுப்படுத்துகிறது. காலப்போக்கில் தமிழ் எழுத்து அடையாளத்தை இழந்துவிட்டு பேச்சு மொழியாக மட்டும் சுருங்கிவிடும் விபத்து நிகழாது என்பதற்குக் காரணங்கள் குறைவாக உள்ளன. தாய்நாட்டிலேயே தமிழன் மொரீசியஸ் தமிழன் ஆகி விடுவானோ என்று அஞ்சுகிறேன். இலங்கைத் தமிழன் நாட்டை இழந்தான். இந்தியத் தமிழன் மொழியை இழந்தான் என்றாகிவிட்டால் செவ்வாய்க் கிரகத்தில் குடியேறுபவனா தமிழ் பேசப் போகிறான்? தாய்மொழி என்பது வயிற்றுப்பாட்டுக்கானதல்ல; பண்பாட்டுக்கானது என்று தமிழினம் புரிந்துகொள்ள வேண்டும். மொழியை இழப்பவன் நிலத்தை இழக்கிறான். நிலம் இழப்பவன் அடையாளம் இழக்கிறான்.”

இருமொழிக் குழந்தைகளுக்கு அதிகத் திறன் மற்றும் படைப்பு சிந்தனை என்று புதிய ஆய்வு

பல்வேறு மக்கள் ஒரு பொதுவான மொழியைப் பகிர்ந்து கொள்வதை ஊக்குவிக்க வலுவான பொருளாதார காரணங்கள் உள்ளன. மொழிபெயர்ப்புக்குச் செலவு அதிகம். இருமொழிக் கொள்கைக்கு ஒரு சிறிய ஆனால் அளவிடக்கூடிய அறிவாற்றல் செலவு (cognitive cost) உள்ளது. இதைத்தவிர கல்விக்கும் அதிக செலவு செய்ய வேண்டும் என்பதெல்லாம் உண்மைதான். இதற்கு மாறாக ஒரு மொழிக்கு மேல் கற்றுக் கொள்ளும் குழந்தைகளுக்கு ஒரு கற்றல் நன்மை உண்டு என்று ஒரு புதிய ஆய்வு வெளிப்படுத்துகிறது. [ஸ்காட்லாந்தில் உள்ள ஸ்ட்ராத்தினைட் பல்கலைக்கழகத்தின் ஆராய்ச்சியாளர்களின்படி](#) ஒரு மொழி மட்டும் தெரிந்த குழந்தைகளைவிட இவர்களுக்குச் சிறந்த சிக்கல் தீர்க்கும் திறன் மற்றும் படைப்புச் சிந்தனை உண்டு. மொழிகளுக்கு இடையில் மாறுவதற்கான மன விரைவியக்கம் மற்றவகை சிந்தனைகளை வளர்ப்பதற்கான திறன்களை உருவாக்கியிருக்கலாம் என்று கருதுகிறார்கள்.

சில புலம்பெயர்ந்த தமிழ் பெற்றோர்கள், தங்கள் பிள்ளைகள் தமிழ் மொழியைக் கற்றுக் கொள்ள உதவுவதற்காக மிகவும் முயற்சி எடுப்பது மகிழ்ச்சி ஊட்டுகிறது. இந்தப் பிள்ளைகள் பெற்றோரின் உதவியுடன் தமிழை இணையத்தில் கற்றுக் கொள்கிறார்கள். அல்லது வார இறுதியில் உள்ளூரில் தன்னார்வலர்கள் நடத்தும் தமிழ் வகுப்புகளுக்குச் செல்கிறார்கள்.

தங்ளிஷ் அல்லது தமிழ் நோயின் மூலகாரணம் என்ன?

நவநாகரிகம் என்று கருதி ஆங்கிலமும் தமிழும் கலந்த தங்ளிஷ் அல்லது தமிழ்நிலம் மொழியில் பேசுவது எழுதுவதைப் பற்றிப் பலர் அச்சிலும் இணையத்திலும் புலம்புகிறார்கள். இவையெல்லாம் நோயின் அறிகுறிகளே. இதற்குப் பதிலாக இந்த நோயின் மூலகாரணத்தைக் கண்டறிவதற்கு இக்கட்டுரைத் தொடரில் முயற்சி செய்வோம். தொடர்ந்து இரண்டு ஆங்கில மொழிப் பேரரசுகள். இவற்றில் ஆங்கிலேயப் பேரரசின் குடியேற்ற நாடாக நம் நாடு ஆயிற்று. நாம் விடுதலை பெற்ற பின்னர் இந்திய அரசியலமைப்பின் எட்டாவது அட்டவணை 22 மொழிகளுக்கு அங்கீகாரம் அளித்துள்ளது. ஆகவே ஆங்கிலம் தொடர்பு மொழியாகத் தேவைப்படுகிறது. பின்னர் உலகமயமாக்கலும், அமெரிக்க கலாச்சார ஆதிக்கமும். மேலும் இப்பெரிய போக்குகள் தகவல் தொழில்நுட்பம் மூலம்

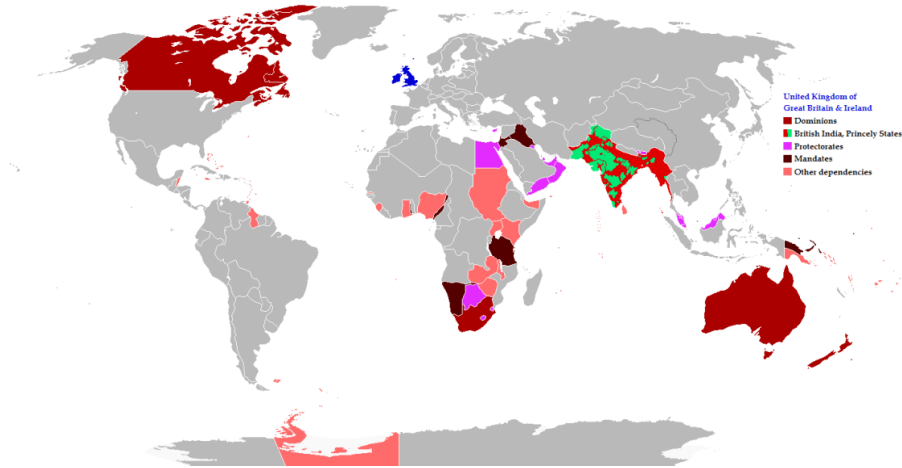
துரிதப்படுத்தப்பட்ட விதம். அதிக சதவீதம் வேலையின்மையும் மற்றும் நிலையற்ற, தகுதிக்குக் குறைவான வேலை நிலையும். வேலை வாய்ப்பை அதிகப்படுத்துவதற்காகப் பெற்றோர்கள் ஆங்கிலக் கல்விக்கு அதீத முக்கியத்துவம் கொடுப்பது. இவைதான் இந்த நோயின் மூலகாரணங்கள் போல் தெரிகின்றன. இவற்றை அடுத்த சில கட்டுரைகளில் ஆழ்ந்து ஆராய்வோம்.

2. தொடர்ந்து இரண்டு ஆங்கில மொழிப் பேரரசுகள்

பேரரசு அல்லது வல்லரசு என்பது மேலாதிக்க நிலைப்பாட்டைக் கொண்ட ஒரு நாட்டை விவரிக்கப் பயன்படுத்தப்படும் சொல். இது உலகளாவிய அளவில் விரிவான முறையில் செல்வாக்கை பயன்படுத்துவது மற்றும் வலிமையைக் காட்டுவதுதான். இது பொருளாதார, இராணுவ, தொழில்நுட்ப மற்றும் பண்பாட்டு வலிமை, அரசியல் செயலாட்சி நயம் மற்றும் செல்வாக்கின் ஒருங்கிணைந்த வழிவகைகளால் செயல்படுகிறது. இரண்டாம் உலகப் போர் முடிவுற்ற காலகட்டத்தில் பிரிட்டிஷ் பேரரசு, அமெரிக்கா மற்றும் சோவியத் ஒன்றியம் மூன்றும் வல்லரசாகக் கருதப்பட்டன.

“பிரிட்டிஷ் பேரரசில் ஞாயிறு மறைவதில்லை”

தொழிற்புரட்சியில் முக்கிய சக்தியாக இருந்ததால் பிரிட்டன் ஒரு சக்திவாய்ந்த பேரரசாக ஆயிற்று. குடிசைத் தொழில், மரபு சார்ந்த விவசாயம் மற்றும் கைவினைகளுக்குப் பதிலாக தொழிற்சாலை அடிப்படையிலான உற்பத்தி, நுட்பமான இயந்திரங்கள், தொழில்நுட்ப வளர்ச்சி, புதிய எரிசக்தி ஆதாரங்கள் மற்றும் போக்குவரத்து முன்னேற்றங்கள் சேர்ந்த பொருளாதார மாற்றங்களைத்தான் தொழிற்புரட்சி என்று சொல்கிறோம். இத்துடன் பல நாடுகளில் குடியேற்றம் வேறு. இவ்வாறு பிரிட்டன் தன்னைப் பேரரசாகக் கட்டி அமைத்ததுதான் ஆங்கில மொழியின் உலகளாவிய பரவலின் முதல் கட்டம். இது ஏகாதிபத்தியம் என்று அழைக்கப்படுகிறது. 1919 ஆம் ஆண்டில் அதன் ஏகாதிபத்தியத்தின் உச்ச கட்டத்தில், பிரிட்டிஷ் பேரரசு வரலாற்றில் மிகப்பெரியது, இது பூமியின் மொத்த நிலப்பகுதியில் சுமார் கால் பங்கு அளவு, 450 மில்லியனுக்கும் அதிகமான மக்கட்தொகை கொண்டது. இதனால்தான் “பிரிட்டிஷ் பேரரசில் ஞாயிறு மறைவதில்லை” என்று பெயர் பெற்றது.



1919 ஆம் ஆண்டில் பிரிட்டிஷ் பேரரசு

17 ஆம் நூற்றாண்டில் கிழக்கு இந்தியா கம்பெனி இந்தியாவில் கொஞ்சம் தொத்த ஆரம்பித்ததிலிருந்து தொடங்கி, 1947 ஆம் ஆண்டு இந்தியா விடுதலை பெற்றதுவரை சுமார் மூன்று நூற்றாண்டுகளாகப் பிரிட்டிஷ் குடியேற்றமாக இருந்தோம். குடியேற்ற நாடுகளில் கல்விக்கு முதன்மை நோக்கம் ஆங்கில மொழியைப் பரப்புவதல். அந்த நாட்டு வாசிகளின் வருங்காலம், கல்வியிலும் வாழ்க்கையிலும் முன்னுக்கு வர முக்கிய வழி அவர்களின் ஆங்கில மொழித் திறன்.

இரண்டாம் உலகப்போரில் பெரும் சேதம் அடைந்ததுடன் பின்னர் குடியேற்ற நாடுகளையும் இழந்திருந்த பிரிட்டிஷ் பேரரசு 1956 இல் சூயஸ் கால்வாய் நெருக்கடியில் பின்வாங்க நேரிட்டதால், ஒரு வல்லரசாக இருந்த செல்வாக்கு பெரிதும் மங்கிவிட்டது.

விருப்பமின்றியே வல்லரசான அமெரிக்கா

இரண்டாவது உலகப் போரிலிருந்து உலகின் முன்னணி இராணுவ மற்றும் பொருளாதார சக்தியாக வெளிவந்த அமெரிக்கா, சோவியத் ஒன்றியத்துடன் மூன்றாம் உலகில் செல்வாக்கிற்காகப் போட்டியிட

வேண்டுமென்று முன் தீர்மானிக்கப்பட்டது போலிருந்தது. அங்கு ஐரோப்பிய ஏகாதிபத்தியத்தின் முடிவுக்குப் பின் ஒரு வலிமை வெற்றிடம் உருவாகியிருந்தது. குறிப்பாக ஆசியா, மத்திய கிழக்கு மற்றும் வட ஆப்பிரிக்காவில் குடியேற்றங்களைக் கட்டுப்படுத்த பிரான்ஸ் மற்றும் பிரிட்டன் போராடியது போலவே அமெரிக்காவும் ஒழுங்கைப் பராமரிப்பதற்கான பொறுப்பை ஏற்றது.

1941 இல் வெளியீட்டு அதிபர் ஹென்றி லூஸ் லைஃப் பத்திரிகையில் “அமெரிக்க நூற்றாண்டு” என்ற தலைப்பில் அமெரிக்காவைப் பற்றிய தனது முன்னோக்குப் பார்வையை ஒரு செல்வாக்குள்ள தலையங்கமாக எழுதினார். அமெரிக்காவை “மேலாதிக்க உலக சக்தியாக”, “உலகம் முழுவதும் கப்பல் போக்குவரத்துக்கு முக்கிய உத்தரவாதமளிப்பவர்” மற்றும் “உலக வர்த்தகத்தின் ஆற்றல் மிகுந்த தலைவர்” என்று விவரித்தார். 1949 இல் வரலாற்றாசிரியர் ஆர்தர் ஷ்லேசிஞ்சர் “உலகத்துக்குத் தலைமை ஏற்கும் பொறுப்பு அமெரிக்காவின் மீது திணிக்கப்பட்டுள்ளது” என்று அறிவித்தார்.

பனிப்போர் முடிவில் அமெரிக்கா உலகின் ஒரே வல்லரசாகியது

நமக்கு விடுதலை கிடைத்த 1947 ஆம் ஆண்டில் தான் ட்ரூமன் கோட்பாடு அறிவிக்கப்பட்டது. இந்த ட்ரூமன் கோட்பாடு என்பது சோவியத் விரிவாக்கத்தால் அச்சுறுத்தப்பட்ட நாடுகளுக்கு உதவுவதற்கான ஒரு அமெரிக்க வெளியுறவுக் கொள்கையாகும். இதில்தான் அமெரிக்காவுக்கும் சோவியத் ஒன்றியத்துக்கும் பனிப்போர் துவக்கம் என்று சொல்லலாம். இந்தப் பனிப்போர்தான் இந்த இரண்டு வல்லரசு நாடுகளும் தமது இராணுவம், தொழில்நுட்பம், மற்றும் விண்வெளி திட்டங்களைப் போட்டிபோட்டுக்கொண்டு வளர்ச்சி செய்ய மூலகாரணமாக இருந்தது. 1989 ஆம் ஆண்டு, கிழக்கு ஐரோப்பாவில் பொதுவுடைமை வீழ்ச்சியுற்று, பெர்லின் சுவரும் உடைந்தது. தொடர்ந்து 1991 ஆம் ஆண்டு சோவியத் ஒன்றியமும் வீழ்ச்சியுற்றது.

அமெரிக்காவின் பொருளாதார, அரசியல் மற்றும் கலாச்சார செல்வாக்கு வளர்ச்சி

அமெரிக்காவின் கலாச்சார, அரசியல் மற்றும் பொருளாதார செல்வாக்கின் உலகளாவிய பரவல் 20 ஆம் நூற்றாண்டில் ஆங்கில மொழியின் மேலாதிக்க நிலையை மேம்படுத்துவதற்கு கணிசமாகப் பங்களித்தது. ஆங்கிலம் இன்று ஐக்கிய நாடுகள் சபை மற்றும் பன்னாட்டு ஒலிம்பிக் கமிட்டியின் அதிகார பூர்வமான மொழியாகும். அறிவியலில் ஆங்கிலமே முழுமையாக ஆதிக்கம் செலுத்துகிறது. 1997 ஆம் ஆண்டில், அறிவியல் மேற்கோள் புள்ளிவிபரம் படி, 95 சதவீத கட்டுரைகள் ஆங்கிலத்தில் எழுதப்பட்டிருக்கின்றன. அவற்றின் ஆசிரியர்களில் பாதி பேர் தான் ஆங்கிலம் பேசும் நாடுகளின் வாசிகள். புள்ளிவிவரங்களின்படி தற்போது உலகெங்கிலும் 60 க்கும் மேற்பட்ட நாடுகள் ஆங்கிலத்தை தங்கள் அதிகார பூர்வ மொழியாகப் பயன்படுத்துகின்றன.

அமெரிக்க தேசிய பாதுகாப்பு ஆலோசகர் ஹென்றி கிசிஞ்சர் 1971 ஜூலையில் பாகிஸ்தானுக்கு பயணம் மேற்கொண்டபோது, ரகசியமாகப் பெய்ஜிங் சென்று அமெரிக்க அதிபர் நிக்சனின் 1972 அதிகாரப்பூர்வமான வருகைக்கு வழி கோலினார். 25 ஆண்டுகளாக எந்த விதத் தொடர்பும் இல்லாதிருந்த சீனாவுடன் உறவைச் சரிசெய்ய அமெரிக்கா விரும்பியது. சோவியத் ஒன்றியத்துடனான உறவுகளில் அதிக செல்வாக்கு பெறவே இந்த யுக்தி. 1990 களில் சோவியத் ஒன்றியம் வீழ்ச்சிக்குப் பின் அமெரிக்காவின் வல்லரசு நிலைமையைச் சவால் செய்ய ஒருவருமில்லை. எனினும் அமெரிக்காவின் தொழிற்சாலை தயாரிப்புகள் வழங்குநராகச் சீனா தன்னை உருவாக்கிக்கொண்டது. ரேடாருக்குக் கீழேயே பறந்து, கண்ணுக்குத்தெரியாமலே பொருளாதார மற்றும் ராணுவ வல்லரசு நிலைமைக்குச் சீனா வளர்ந்து விட்டதைப் பார்க்கிறோம்.

அடுத்து நிகழவிருக்கும் சீன பொருளாதார மற்றும் இராணுவ ஆதிக்கத்திலும் கவனம் வைப்போம்

சீனாவின் பொருளாதாரம் 2035 க்குள் அமெரிக்காவை விட அதிகமாக இருக்குமெனச் சில கணிப்பாட்டிகள் தெரிவிக்கின்றன. சீனாவின் விரைவான வளர்ச்சி உள்நாட்டுத் தேவைகளால் உந்தப்படுகிறது. முன் போல் ஏற்றுமதியினால் அல்ல. ஆகவே இந்த நூற்றாண்டில் பெரும்பாலும் தொடர்ந்து சுமார் 6 முதல் 7 விழுக்காடு வரை வளர்ச்சி நீடிக்கும் என்றும் சொல்கிறார்கள். அமெரிக்காவோ ஏற்கெனவே வளர்ச்சிபெற்ற நாடு. ஆகையால் அதன் வளர்ச்சி சுமார் 2 முதல் 3 விழுக்காடுதான்.

உலகின் முதன்மை பொருளாதார சக்தி என்ற இடத்துக்கு அமெரிக்காவுக்கு சவால் விடச் சீனா வழிசெய்து கொண்டிருக்கிறது. சீன ராணுவமும் பல நாடுகளுடன் எல்லைப் பிரச்சினை மற்றும் நிலம், கடல்பகுதி விவகாரங்களில் பலாத்காரமாகத் தலையிடுவதையும் பார்க்கிறோம். அருணாசல பிரதேசத்தைத் தெற்கு திபெத் என்று உரிமை கொண்டாடி பிரச்சினை எழுப்பிக் கொண்டிருப்பதையும் பார்க்கிறோம். தகவல் தொழில்நுட்பத்திலும் சீனா தனிவழி அமைத்துக் கொண்டது மட்டுமல்லாமல் சில துறைகளில் அமெரிக்காவை மிஞ்சும் நிலையில் இருப்பதாகச் சொல்கிறார்கள். ஆனால் இதன் விளைவாக மொழியியல் மேலாதிக்கம் ஏற்படுமா என்பதுதான் கேள்வி.

“அரசியல், பொருளாதாரம், பண்பாடு, தொழில்நுட்பம் அல்லது இராணுவம் தனியாக ஒரு மொழிக்குப் பன்னாட்டு முக்கியத்துவத்தை வழங்க முடியாது. இதை அடைவதற்கு இந்த எல்லா சக்திகளும் ஒன்று சேர்ந்த தொடர்ச்சியான தாக்கம் இருக்க வேண்டும். 1960 களிலிருந்து 1990 கள் வரை ஜப்பான் நம்பமுடியாத பொருளாதார வெற்றியைக் கண்டிருந்தபோதிலும், ஜப்பானிய மொழி பெரு வழக்கு அடைந்தது, ஆனால் பன்னாட்டு அளவில் ஒரு முக்கிய மொழியாக வர இயலவில்லை.” என்று ஒரு கருத்து நிலவுகிறது. ஆனால் ஜப்பானிய மொழி பேசுவோர் சுமார் 150 மில்லியனுக்குக் கீழேதான். சீன மொழி பேசுவோரோ ஹாங்காங், தைவான், சிங்கப்பூர் உட்பட 1.3 பில்லியன். ஜப்பானில் மக்களாட்சி, சீனாவிலோ கொடுங்கோன்மை. சீனாவில் வாழும் திபெத், வீகர் இன மக்களைத் தங்கள் தாய் மொழியைக் கற்க வழி இல்லாமல் செய்து, சீன மொழியைக் கற்கக் கட்டாயப்படுத்துவதையும் மறந்து விடக் கூடாது.

3. உலகமயமாக்கலும் தகவல் தொழில்நுட்பமும்

கோட்பாட்டைப் பொருத்தவரை உலகமயமாக்கல் நன்றாகத் தானிருக்கிறது. உங்கள் நாட்டில் எந்தப் பொருட்களைக் குறைந்த செலவில் நல்ல தரத்தில் செய்ய முடியுமோ அவற்றை உலகெங்கிலும் ஏற்றுமதி செய்யுங்கள். அந்த வருமானத்தை வைத்து உங்கள் நாட்டுக்குத் தேவையான பொருட்களை அவற்றைத் திறமையாகச் செய்யும் நாடுகளிடமிருந்து இறக்குமதி செய்து கொள்ளுங்கள்.

இம்மாதிரி தன் கையே தனக்குதவி என்று இல்லாமல் அன்றாட இன்றியமையாத பொருட்களுக்குக் கூட மற்ற நாடுகளை நம்பி இருப்பதுதான் உலகமயமாக்கல். இதனுடைய ஒரு துணை விளைவுதான் ஆங்கில மொழி ஆதிக்கம். தகவல் தொழில் நுட்பத்துறையில் உலகமயமாக்கலின் தாக்கம் பரவலாகத் தெரிந்த சங்கதி. ஆகவே தயாரிப்பு தொழிற்சாலைத் துறையில் ஒரு எடுத்துக்காட்டை ஆய்வு செய்வோம்.

உலகமயமாக்கலுக்கு நடைமுறை எடுத்துக்காட்டு – திருப்பூர்

முப்பது ஆண்டுகளில் திருப்பூர் பின்னிய உள்ளங்கிகளின் தலைநகரம் ஆகிவிட்டது. “டாலர் நகரம்” அல்லது “சிறிய ஜப்பான்” என்று பரவலாக அழைக்கப்படும் இந்நகரம் பின்னிய ஆயத்த ஆடைகளில் சிறந்து விளங்குகிறது. 1970 களில், ஒரு இத்தாலியக் கொத்தணியுடன் இணைந்து திருப்பூர் ஏற்றுமதி சந்தையில் வாய்ப்புகளைப் பயன்படுத்தத் தொடங்கியது. பின்னர் ஐரோப்பா, அமெரிக்கா மற்றும் பசிபிக் நாடுகளில் நல்ல பெயர் பெற்று இன்று திருப்பூர் ஒரு முன்னிலை ஏற்றுமதிக் கொத்தணியாக விளங்குகிறது.

1984 இல் ஏற்றுமதி 10 கோடி ரூபாய். சுமார் இருபது ஆண்டுகள் பின்னர் 2006-07 ஆம் ஆண்டில் ஏற்றுமதி 11,000 கோடி ரூபாய். ஏற்றுமதி மூலம் 3.5 லட்சம் மக்களுக்குத் திருப்பூர் வேலைவாய்ப்பு அளிக்கிறது. இவ்வளவு குடும்பங்களுக்கு வாழ்வாதாரமாகவுள்ள உலகமயமாக்கலை வேண்டாம் என்று சொல்ல முடியுமா?

நீங்கள் தயாராக இருக்கிறீர்களோ இல்லையோ, ஆங்கிலம்தான் இப்போது வணிக உலக மொழி

சில கணிப்பீடுகளின்படி ஆங்கிலம் நன்றாகத் தெரிந்த நாடுகளின் பொருளாதாரங்கள் சிறப்பாக உள்ளன மக்கள்தொகையின் ஆங்கிலத் திறன்களுக்கும் நாட்டின் பொருளாதார செயல்திறனுக்கும் இடையே நேரடி தொடர்பு உள்ளதாக ஆராய்ச்சிகள் காட்டுகின்றன. மற்றும் தனிப்பட்ட அளவில், உலகெங்கிலும் தங்கள் நாட்டினுடைய சராசரியுடன் ஒப்பிடுகையில் ஆங்கிலத்தில் அதிக திறமையுள்ள வேலை தேடுபவர்கள் 30-50% அதிக சம்பளத்தை பெறுகின்றனர் என்று ஆளெடுப்பவர்களும் மனிதவள மேலாளர்களும் கூறுகிறார்கள். வருமானம் மட்டுமல்ல, வாழ்க்கை தரத்தையே மேம்படச் செய்கிறதாம். ஆங்கிலத்திறமை மற்றும் மனித வளர்ச்சி குறியீட்டிற்கும் இடையேயான ஒரு தொடர்பையும் கண்டறிந்ததாகச் சொல்கிறார்கள். இது கல்வி, ஆயுள் எதிர்பார்ப்பு, கல்வியறிவு மற்றும் வாழ்க்கைத் தரங்களின் அளவு. ஆகவே உலகெங்கிலும் உள்ள பில்லியன் கணக்கான மக்கள், ஆங்கில மொழியைக் கற்றுக்கொள்ளத் தீவிரமாக முயற்சி செய்கிறார்கள். தங்கள் முன்னேற்றத்திற்காக மட்டுமல்ல, இது ஒரு இன்றியமையாத பொருளாதாரத் தேவையாகவும் இருக்கிறது.

உலகமயமாக்கலின் விளைவாக ஆங்கில மொழி ஆதிக்கம்: நடைமுறை எடுத்துக்காட்டு – ரகுடென்

ரகுடென் (Rakuten) என்பது ஜப்பான் நாட்டின் மிகப்பெரிய மின் வணிக நிறுவனம். அதன் தலைமை நிர்வாக அதிகாரி ஹிரோஷி மிக்கிடானி (Hiroshi Mikitani) 2010 இல் ஆங்கிலமே நிறுவனத்தின் அதிகாரபூர்வ மொழி என்று உத்தரவிட்டார். இந்த நிறுவனத்தின் இலக்கானது, உலகின் முதல் இணைய சேவை நிறுவனமாக மாறுவது. இந்தப் புதிய கொள்கை 7,100 ஜப்பானிய ஊழியர்களைப் பாதித்தது. குறிப்பாக ஜப்பானுக்கு வெளியே மற்ற நாடுகளில் விரிவாக்கத் திட்டங்கள் இருந்ததால், ஆங்கிலம் இந்த இலக்கை அடைவதற்கு மிகவும் முக்கியம் என்று மிக்கிடானி நம்பினார்.

மொழி மாற்றம்பற்றித் தீவிரமாக இருந்த மிக்கிடானி திட்டத்தை அறிவித்ததே ஆங்கிலத்தில்தான். அன்றிரவே உணவு விடுதியில் உணவுப் பட்டியல் முதல் மின்தூக்கியில் பெயர்த்தொகுதி வரை எல்லாமே ஆங்கிலமயமாயின. மேலும் இரண்டு ஆண்டுகளுக்குள் பன்னாட்டு ஆங்கில மதிப்பீட்டின்படி பணியாளர்கள் அனைவரும் தகுதிப்பட வேண்டும் அல்லது பதவி இறக்குதல் அல்லது பணிநீக்கம் ஆக நேரிடும் என்றும் அவர் அறிவித்தார். மிக்கிடானியின் அறிவிப்பு ஊழியர்களுக்கு அதிர்ச்சியை அளித்தது. இதுநாள்வரை அவர்கள் ஜப்பானிய மொழியிலேயே வேலை செய்து வந்தனர். அவர்களின் அனைத்து ஆவணங்களும் ஜப்பானிய மொழியில் இருந்தன. ஆங்கிலத்தில் எல்லாவற்றையும் அவர்களால் எப்படிச் சமாளிக்க முடியும்?

பதினெட்டு மாதங்களுக்குப் பிறகு முன்னேற்றத்தை நிறுவனம் மதிப்பீடு செய்தது. சில ஊழியர்கள் கூட ஆங்கிலத்தில் திறமையான அளவை அடையவில்லை. பிரச்சினை என்னவென்றால் பணியாளர்கள், பதிவு செய்த பாடங்களைக் கேட்டு அல்லது புத்தகங்களைத் தங்கள் சொந்த நேரங்களில் படிப்பதன் மூலம், தாங்களே கற்றுக் கொள்ள வேண்டும் என நிறுவனம் எதிர்பார்த்ததுதான்.

அதைப் பார்த்து மிக்கிட்டானி தனது மனதை மாற்றிக்கொண்டார். பின்னர் நிறுவனச் செலவில் ஆங்கில வகுப்புகள், இணையக்கற்றல், செயலிகள் மற்றும் தனிப்பட்ட முறையில் பயிற்சி போன்ற பல பயிற்சிகளைக் கொடுத்து 2015 ஆம் ஆண்டளவில் பெரும்பான்மையான ஜப்பானிய ஊழியர்கள் குழுசந்திப்பில் என்ன நடக்கிறது என்பதை ஓரளவு புரிந்து கொள்ள முடியும் என்ற நம்பிக்கை நிலைக்கு வந்தனர். “என் சக ஊழியர்களுடன் ஆங்கிலத்தில் பேசுவதில் எனக்குப் பயமில்லை,” என்று ஒருவர் கூறினார்.

இம்மாதிரி மனித வரலாற்றில் மிக வேகமாகப் பரவிவரும் மொழி ஆங்கிலம். அமெரிக்கா மற்றும் ஆஸ்திரேலியா போன்ற நாடுகளில் ஆங்கிலத்தை தாய் மொழியாகக் கொண்டவர்கள் 385 மில்லியன் உள்ளனர். இந்தியா மற்றும் நைஜீரியா போன்ற முன்னாள் குடியேற்ற நாடுகளில் ஒரு பில்லியன் சரளமாக ஆங்கிலம் பேசுபவர்கள். உலகம் முழுவதிலும் ஆங்கிலத்தை இரண்டாவது மொழியாகப் படித்தவர்கள் மில்லியன் கணக்கான உள்ளனர். ஆக உலகளவில் சுமார் 1.75 பில்லியன் மக்களால் இது ஒரு பயனுள்ள அளவில் பேசப்படுகிறது.

உலகமயமாக்கலுக்கு ஊக்க மருந்து தகவல் தொழில்நுட்பம்

இணையமும் மற்ற பல தகவல் தொழில்நுட்பங்களும் வளர்ந்தது உலகமயமாக்கலுக்கு ஊக்க மருந்து கொடுத்தாற்போல ஆயிற்று. 1990 களின் முற்பகுதியில் கணினி வன்பொருள், மென்பொருள், மற்றும் தகவல்தொடர்பு மேம்பாடுகள் மக்களுக்குப் பொருளாதாரத் திறனை அணுகும் ஆற்றலை அதிகரித்தன. அதன் பின்னர் இணையத்தின் இரண்டாவது முன்னேற்றத்தில் வந்த சமூக ஊடகங்கள் இணைய அடிப்படையிலான கருவிகளில் தனிப்பட்ட, அரசியல் மற்றும் வர்த்தக நோக்கங்களுக்காக மக்கள் தகவலைப் பயன்படுத்தும், பகிரும் வழிகளையே மாற்றியமைத்தன.

உலகத்தில் எந்த இடம் என்று பொருட்படுத்தாமல், புதுமையான வழிகளில் வளங்களைப் பயன்படுத்தி நாடுகள் மற்றும் கலாச்சாரங்களிடையில் புதிய தயாரிப்புகள் மற்றும் சேவைகளை உருவாக்கத் தகவல் தொழில்நுட்பம் வழி செய்கிறது. தகவல் பரிமாறத் திறமையான மற்றும் பயனுள்ள அலைத்தடங்களை உருவாக்கி, தகவல் தொழில்நுட்பத்துறை உலகளாவிய ஒருங்கிணைப்பு ஊக்கியாக இருக்கிறது.

4. அமெரிக்க கலாச்சார ஆதிக்கமும் தகவல் தொழில்நுட்பமும்

முதலில் ஹாலிவுட் திரைப்படங்களும் தொலைக்காட்சி நாடகத் தொடர்களும்

1920 களில் தொடங்கி அமெரிக்க திரைப்படம் மற்றும் தொலைக்காட்சித் தொழில் மையமான ஹாலிவுட், உலகின் பெரும்பாலான ஊடகச் சந்தைகளில் ஆதிக்கம் செலுத்தி வருகிறது. உலகெங்கிலும் உள்ள மக்கள் அமெரிக்க பாணி, பழக்கவழக்கம் மற்றும் வாழ்க்கை முறை ஆகியவற்றைப் பற்றித் தெரிந்து கொள்ளவும் அவற்றைப் பின்பற்ற முயற்சி செய்யவும் முக்கிய ஊடகம் இது.

ஹாலிவுட் திரைப்படங்கள் ஆங்கில மொழியைப் பரப்புகின்றன மற்றும் பிரபலப்படுத்துகின்றன. ஆங்கில மொழி உலக மொழியாக மாறி, உலகின் தலைமையைத் தக்கவைத்துக்கொள்வதற்கு முக்கியக் காரணமாகவும் உள்ளது. இத்திரைப்படங்களைப் பார்த்தே தினசரி வாழ்க்கையில் பேசும் ஆங்கில சொற்றொடர்களைப் பேச மக்கள் எளிதாகப் பழகி விடுகின்றனர்.

அடுத்து கொகா கோலா, பெப்சி, மெக்டொனால்ட்ஸ், பர்கர் கிங், பிட்சாஹட், டாமினோஸ், கேள்ப்சி முதலியன

உலகின் பிரபலமான முதல் பத்து வணிகச் சின்னங்களில் ஏழு அமெரிக்காவைச் சேர்ந்தவை. இவற்றில் முதலிடத்தில் உள்ள கொகா கோலா அமெரிக்கமயமாக்கல் சின்னமாகவே கருதப்படுகிறது. துரித உணவு கூட அமெரிக்க உணவு சந்தையிலிடுதல் மேலாதிக்கத்திற்கான ஒரு சின்னமாகவும் கருதப்படுகிறது. மெக்டொனால்ட்ஸ், பர்கர் கிங், பிட்சா ஹட், கேள்ப்சி மற்றும் டாமினோஸ் பிட்சா போன்ற நிறுவனங்களின் துரித உணவு விடுதிகள் உலகெங்கிலும் ஏராளமான உள்ளன.



பின்னர் மைக்கேல் ஜாக்ஸன் இசை, ஜீன்ஸ், டைட்ஸ், கிழிந்த ஜீன்ஸ் இன்னபிற

எல்விஸ் பிரெஸ்லி மற்றும் மைக்கேல் ஜாக்ஸன் போன்ற பல அமெரிக்க கலைஞர்களும் உலகளவில் ரசிக்கப்பட்டு 500 மில்லியன் தொகுப்புகளை விற்றுள்ளனர். த்ரில்லர் என்ற மைக்கேல் ஜாக்ஸனின் தொகுப்பு, 100 மில்லியன் விற்பனையில், எல்லாக் காலத்திலும் சிறந்த விற்பனையான தொகுப்பாக உள்ளது.

1989-91 ஆம் ஆண்டில் சோவியத் ஒன்றியம் விழுந்தபின் வலிவடைந்து, குறிப்பாக 2000 ஆம் ஆண்டுகளின் மத்தியில் வந்த அதிவேக இணையத் தொடர்பின் பின்னர், அமெரிக்கமயமாக்கல் மிகவும் பரவலாகிவிட்டது.

புதுக்குடியேற்றம்

1960 களில் பல ஆப்பிரிக்க நாடுகளுக்கு குடியேற்றத்திலிருந்து விடுதலை கிடைத்தது. இந்த சூழலில், நேரடியாக இராணுவக் கட்டுப்பாட்டிலோ (imperialism) அல்லது மறைமுகமான அரசியல்

கட்டுப்பாட்டிலோ (hegemony) வைப்பதற்குப் பதிலாக உலகமயமாக்கல், கலாச்சார மற்றும் மொழி ஆதிக்கத்தின் செல்வாக்கைப் பயன்படுத்தி இம்மாதிரி வளர்ந்து வரும் நாடுகளில் காரியம் சாதித்துக்கொள்வதை புதுக்குடியேற்றம் என்று கானா நாட்டின் தலைவர் குவாமே நிக்ரூமா வருணித்தார். மேற்கத்திய சிந்தனையாளர்களான ஜீன்-பால் சார்த்ரே மற்றும் நோம் சோம்ஸ்கி போன்றவர்களின் படைப்புகளிலும் கூட இந்தப் புதுக்குடியேற்றப் போக்கு விவாதிக்கப்பட்டது.

இன்றைய உலகளாவிய கலாச்சாரம் பன்முக கலாச்சாரத்தைப் போல் தோற்றமளித்த போதிலும், கடைசியில் பகுப்பாய்வு செய்து பார்த்தால் அமெரிக்க ஒரே வகைக் கலாச்சாரமாகவுள்ளது என்று ஜப்பானிய அறிஞர் டோரு நிஷிகாகி (Toru Nishigaki) வாதிடுகிறார். இந்த அமெரிக்கக் கொள்ளை நோய் மற்ற அனைத்துக் கலாச்சாரங்களையும் தொற்று நோய் போல் பிடிக்கவும், பாதிக்கவும், இகழவும் மற்றும் தள்ளிவைக்கவும் அச்சுறுத்துகிறது என்று கூறுகிறார்.

அமெரிக்க கலாச்சார ஆதிக்கத்துக்கு ஊக்க மருந்து தகவல் தொழில்நுட்பம்

மைக்ரோசாப்ட், ஆப்பிள், இன்டெல், ஹெச்பி, டெல் மற்றும் ஐபிஎம் போன்ற உலகின் மிகப் பெரிய கணினி நிறுவனங்கள் அமெரிக்காவைச் சேர்ந்தவை. மேலும் உலகளாவிய ரீதியில் வாங்கப்படும் மென்பொருட்களும் மேகக்கணிமை என்று சொல்லப்படும் இணைய சேவைகளும் அமெரிக்க நிறுவனங்களால் உருவாக்கப்பட்டவை. ஆக உலகளாவிய அடிப்படையில், குறிப்பிட்டு மென்பொருள் துறையிலும் பொதுவாக தகவல் தொழில்நுட்பத் துறையிலும் அமெரிக்கா மிக வலுவான நிலையில் உள்ளது.

உலகின் மிகப்பெரிய தொழில்நுட்ப நிறுவனங்களான பேஸ்புக், நெட்ஸ்பைக்ஸ் மற்றும் கூகிள் போன்ற சிலவற்றிற்கு அமெரிக்காவில் கலிபோர்னியாவிலுள்ள சிலிக்கான் பள்ளத்தாக்கு தொழில்நுட்ப தலைநகரமாகவும் உலகிலேயே மிக வெற்றிகரமான தொடக்கநிலை சுற்றுச்சூழல் அமைப்பாகவும் அமைந்துள்ளது. திறமையான பணியாளர்கள், மூலதனம் மற்றும் பெரிய கருத்துக்கள் ஒரே இடத்தில் கூடி வருவதால், தொடக்கநிலை நிறுவனங்கள் வளர இது சிறந்த இடமாகிறது. சமீபத்திய ஆண்டுகளில் கூகிள், முகநூல், ட்விட்டர், அமேசான், ஆப்பிள் மற்றும் ஊபர் போன்ற அமெரிக்க இணைய நிறுவனங்கள் மூலம் அமெரிக்கமயமாக்கல் குறித்த கவலை ஐரோப்பாவில் அதிகரித்து வருகிறது.

இந்த உலகளாவிய பெரும் போக்கில் மாட்டாமல் பைடு தேடல் இயந்திரம், வெய்போ சமூக ஊடகம், வீசேட் அரட்டை, டிடி வாடகை வண்டி, அலிபாபா, டென்சென்ட் போன்ற பல இணைய தகவல் தொழில்நுட்ப சேவைகளையும், நிறுவனங்களையும் உருவாக்கி சீனா தனக்கென தனி வழி அமைத்துக் கொண்டதும் குறிப்பிடத்தக்கது.

இந்தியை ஆட்சி மொழியாக்கி விடலாமே என்று சிலர்

சீனா செல்லும் இந்தத் தனி வழியைப் பார்த்து வடக்கில் உள்ள கட்சிகளில் சிலர் இந்தியாவில் இந்தியை ஆட்சி மொழியாக்கி விட்டால் நாமும் சீனா மாதிரிச் செய்யலாமே என்று நினைக்கிறார்கள். இந்திய அரசியலமைப்பின் எட்டாவது அட்டவணைப்படி இந்தியாவின் அதிகாரபூர்வ மொழியாக 22 மொழிகள் உள்ளன. இந்தி அவற்றில் ஒன்றுதான்.

புதிய பாதுகாப்பு மந்திரியுடன் அவர்களின் முதல் சந்திப்புக்கு வரும் போது, மூன்று இராணுவத் தலைவர்களையும் (தரைப்படை, கடற்படை, வான்ப்படை) நீங்கள் இந்தியர்கள் என்பதை நினைவுபடுத்துவது அரிது. ஆனால் 1996 இல் பாதுகாப்பு மந்திரியான முலாயம் சிங் யாதவ் ‘நீங்கள் இந்தியர்கள், எனவே இந்தியில் பேசுங்கள்,’ என்று கூறினார். உடன் அவர் ஒரு உத்தரவையும் வெளியிட்டார், அனைத்து அதிகாரபூர்வமான அறிவிப்புகளும் இந்தியில் இருக்க வேண்டும் என்று. “இது நாட்டின் தேசிய மொழி என்பதால் இந்தியாவின் குடிமக்கள் இந்தி கற்றுக்கொள்ள முயற்சிக்க வேண்டும்” என்று மத்திய அமைச்சர் எம். வெங்கையா நாயுடு 2017 ஜூன் மாதம் கூறினார். பின்னர் இவர் இந்தியக் குடியரசுத் துணைத் தலைவராகவும் ஆகிவிட்டார்.

இந்த ஆண்டு 2018 ஜனவரியில் மக்களவையில் பேசுகையில், இந்தியை ஐ.நா.வில் அதிகாரபூர்வ மொழியாக ஆக்க எந்தவிதமான முயற்சியையும் மேற்கொள்ள அரசாங்கம் தயாராக உள்ளதாக

வெளியுறவு அமைச்சர் சுஷ்மா ஸ்வராஜ் வலியுறுத்தினாராம். இந்தி தேசிய மொழி அல்ல, இது ஒரு அதிகாரபூர்வ மொழிதான். இந்தி பேசாத மாநிலங்களின் பிரதிநிதிகள் எதிர்காலத்தில் நாட்டின் தலைவர்களாக வரக்கூடும். அவர்கள் மீது இந்தி மொழியைச் சமத்த வேண்டிய அவசியமில்லை என்று எதிர்க்கட்சித் தலைவர் சசி தாநூர் சுட்டிக் காட்ட வேண்டியது அவசியமாயிற்றாம். இவர்கள் யாவரும் சட்டப்படி இந்தியாவின் அரசியலமைப்பில் உண்மையான பற்றுதல் மற்றும் நம்பிக்கை வைப்பதாக (true faith and allegiance) உறுதிமொழி எடுத்துக்கொண்டவர்கள்.

ஐரோப்பிய ஒன்றியம், கனடாவின் சிறுபான்மை மொழிக் கொள்கைகளைப் பார்த்துக் கற்றுக் கொள்ளலாமே

ஐரோப்பிய ஒன்றியத்தின் அன்றாட நடவடிக்கைகளில் மொழிகள் முக்கிய பாத்திரம் வகிப்பது மட்டுமல்ல இவை ஐரோப்பிய ஒன்றியத்தின் கலாச்சார மற்றும் மொழியியல் பன்முகத்தன்மைக்கும் அடிப்படையாகும். மொழி பன்முகத்தன்மையும் மற்ற கலாச்சாரங்களுக்கு மரியாதை என்பதும் ஐரோப்பிய ஒன்றியத்தின் அடிப்படை மதிப்பாகும். ஒவ்வொரு குடிமகனுக்கும் ஐரோப்பிய ஒன்றியத்தின் எந்தப் பிரிவுக்கோ அல்லது துறைக்கோ எந்தவொரு அதிகாரபூர்வமான மொழியிலும் எழுதவும் அதே மொழியில் பதிலைப் பெறவும் உரிமை உண்டு.

ஐரோப்பிய பாராளுமன்றம் ஒரு முழு பன்மொழிக் கொள்கையை ஏற்றுள்ளது, அதாவது அனைத்து ஐரோப்பிய ஒன்றிய மொழிகளும் சமமாக முக்கியத்துவம் வாய்ந்தவை. அனைத்து நாடாளுமன்ற ஆவணங்களும் அனைத்து அதிகாரபூர்வ மொழிகளிலும் மொழிபெயர்க்கப்படுகின்றன. ஐரோப்பிய பாராளுமன்றத்தின் ஒவ்வொரு உறுப்பினருக்கும் அவர் விரும்பிய அதிகாரபூர்வ மொழியில் பேசும் உரிமை உண்டு.

கனடா நாட்டிலும் பிரெஞ்சு மொழி பேசுவோர் சிறுபான்மை என்ற போதிலும் அவர்களுக்கும் பிரெஞ்சு மொழிக்கும் சம உரிமை உண்டு. நம் நாடாளுமன்ற உறுப்பினர்களுக்கும் அமைச்சர்களுக்கும் இந்த அடிப்படைக் கண்ணியம் உள்ளதென்றும் மற்றும் இந்தியாவின் அரசியலமைப்பில் இவர்கள் எடுத்துக்கொண்ட உறுதிமொழியைக் காப்பாற்றுவார்கள் என்றும் நம்பிக்கை வைப்போம்.

5. பரவும் இந்த ஆங்கில மொழி ஆதிக்கம் தணிய வாய்ப்பு உள்ளதா?

ஆங்கிலம் மற்ற மொழிகளையும் கலாச்சாரங்களையும் அழிக்கிறதா?

உலகமயமாக்கல், அமெரிக்க கலாச்சாரம் மற்றும் தொழில்நுட்பப் பேரலைகளின் மேல் ஏறி உலகில் இதுவரை எந்த மொழிக்கும் இல்லாத அளவுக்கு ஆங்கிலம் ஆதிக்கம் செலுத்துகிறது. சீன அல்லது ஸ்பானிஷ் மொழியைத் தாய்மொழியாகக் கொண்டவர்கள் ஆங்கிலத்தைத் தாய்மொழியாகக் கொண்டவர்களை விட அதிகமாக இருக்கலாம், ஆனால் வேற்று மொழியாளர்களுடன் பேசும் போது அவர்களும் ஆங்கிலம்தான் பேச வேண்டியிருக்கிறது. மேலும் பிணைப்பு பெருகி வரும் உலகின் குடிமக்களாக ஆக்குவதற்காக அவர்களும் தங்கள் குழந்தைகளுக்கு ஆங்கிலம்தான் கற்றுக்கொடுக்கிறார்கள். “அனைவருக்கும் ஆங்கிலம் இரண்டாம் மொழியாகி விட்டது” என்று தலிபோர்னியாவில் இர்வின் பல்கலைக்கழகத்தின் கல்வி மற்றும் தகவல் தொடர்புத்துறை பேராசிரியரான மார்க் வார்ஷூவர் கூறுகிறார். “உலகின் எந்தப் பகுதியாக இருந்தாலும் கல்வி பயின்றவர் என்றால் ஆங்கிலம் தெரிந்தவர் என்ற நிலை வந்துவிட்டது.”

ஆங்கில ஆதிக்கமும் உலகமயமாக்கலும் ஒன்றன் பின் ஒன்றாக உலகம் முழுவதும் பரவின. “உலகளாவிய மொழி ஒன்று இருப்பது உலகமயமாக்கலுக்கு உதவியது, உலகமயமாக்கல் உலகளாவிய மொழியை ஒருங்கிணைத்தது,” என்று வார்ஷூவர் கூறுகிறார். தொடர்ச்சியாக ஆங்கில மொழி பேசும் இரண்டு பேரரசுகள், ஆங்கிலேயர்கள் மற்றும் அமெரிக்கர்களின் ஆதிக்கத்தில் அந்தப் போக்குத் தொடங்கி இன்று இணையத்தின் புதிய மெய்நிகர் பேரரசு மூலம் தொடர்கிறது.

பரவலாக ஏற்றுமதி செய்யப்படும் அமெரிக்கத் திரைப்படங்கள், தொலைக்காட்சி நிகழ்ச்சிகள், துரித உணவகங்கள் மற்றும் நுகர் பொருட்கள் ஆகியவை அமெரிக்க கலாச்சார ஆதிக்கத்தை நிலைநாட்டி பிற நாடுகளின் பாரம்பரிய கலாச்சாரங்களை முற்றுகையிடுகின்றன. இதைவிட முக்கியமாக இவை இளைய தலைமுறையினரை எளிதில் வயப்படுத்தத்தக்க வயதில் கவர்ந்து வாழ்நாள் முழுவதும் ஆங்கில மொழி ஆதிக்கத்துக்கு உள்ளாக்குகின்றன. இந்த இருண்டு வரும் ஆங்கில மொழி ஆதிக்கமென்ற மேக மூட்டத்தின் கீழ் மாட்டிக்கொண்ட தமிழ் போன்ற மொழிகளுக்கு சிறு கீற்றொளிக்குக் கூட வழி கிடையாதா?

பிரான்ஸ் எக்காலத்திலும் ஒரு ஆங்கில மொழி பேசும் நாடாக மாறப்போவதில்லை

ஒரு பிரபலமான விவாதக்குழுவில் “ஆங்கிலம் மற்றெல்லா மொழிகளையும் அழிக்கிறதா?” என்ற கேள்விக்கு ஒருவர் கூறுகிறார், “இந்தக் கேள்வியின் பிற பதில்கள் ஆங்கில ஆக்கிரமிப்பின் மற்ற விளைவுகளைச் சுட்டிக்காட்டுகின்றன. எனினும் பிரான்ஸ் எக்காலத்திலும் ஒரு ஆங்கில மொழி பேசும் நாடாக மாற வாய்ப்பேயில்லை. சீனா, ரஷ்யா, தாய்லாந்து போன்ற ஆங்கில மொழி பேசாத மற்ற நாடுகளும் இதேபோல்தான். ஆகவே நாடுகள் இருக்கும் வரை ஆங்கில கலாச்சார ஆதிக்கம் வளரக்கூடும், ஆனால் ஆங்கில மொழி ஆதிக்கம் வளர வாய்ப்பில்லை.”

15 ஆம் நூற்றாண்டில் லத்தீன் மொழிக்கு ஒளிமயமான எதிர்காலம் என்றுதான் எண்ணினார்கள்!

“15 ஆம் நூற்றாண்டு ஐரோப்பாவின் மனநிலையில் இருந்தால் லத்தீன் மொழிக்கு எதிர்காலம் மிகவும் ஒளிமயமாக இருக்கிறது என்றுதான் நினைப்போம்” என்று லத்தீன் வரலாற்றை எழுதிய நிக்கோலஸ் ஓஸ்ட்லர் கூறுகிறார். இன்று கத்தோலிக்க தேவாலயத்தின் வழிபாட்டு மொழியாக மட்டுமே லத்தீன் பிழைத்துள்ளது. “20-ஆம் நூற்றாண்டின் உலகின் மனநிலையில் நீங்கள் இருந்தால், ஆங்கில மொழிக்கு எதிர்காலம் மிகவும் ஒளிமயமாக இருக்கிறது என்று நினைப்பதும் அதேபோல்தான்.” ஆனால் நிக்கோலஸ் ஓஸ்ட்லர் மாதிரி நினைப்பவர்கள் மிகச் சிறுபான்மையிலேயே இருக்கிறார்கள்.

ஆங்கிலத்தை மேம்பட்ட நிலைக்குக் கொண்டுவர உதவிய அதே தொழில்நுட்பமே அதை மீண்டும் கீழே இழுக்கவும் முடியும்

இறுதியில் ஆங்கிலத்தை மேம்பட்ட நிலைக்குக் கொண்டுவர உதவிய அதே தொழில்நுட்பமே அதை மீண்டும் கீழே இழுக்கவும் முடியும் என்று ஓஸ்ட்லெர் கூறினார். கூகிள், மைக்ரோசாப்ட் பிங் மற்ற பெரிய நிறுவனங்களின் மொழிபெயர்ப்பு இயந்திரங்கள் ஏற்கெனவே ஓரளவு நன்றாக வேலை செய்கின்றன. இறுதியில் ஒரு அதி உன்னத மொழிபெயர்ப்பு இயந்திரத்தை உருவாக்கி விட்டால் ஒரு பொதுவான மொழி தேவையற்றுது என்று ஒரு கருத்தும் நிலவுகிறது. “முன்னேற்றம் வர வர, இயந்திர மொழிபெயர்ப்பு மற்றும் தானியங்கி உரை விளக்கம் பற்றிய பிரச்சினைகள் தீர்ந்துவிடும்,” என்று ஓஸ்ட்லெர் கூறினார், “தொழில்நுட்ப ரீதியாக ஒரு பொதுவான மொழி தேவையற்றுப் போய்விடும்.”

நீங்கள் ஒரு திருப்பூர் பின்னிய உள்ளங்கி வழங்குநர் என்று வைத்துக்கொள்வோம். உங்கள் திறன்பேசியை எடுத்து ஜெர்மனியில் உள்ள உங்கள் வாடிக்கையாளரைக் கூப்பிடுகிறீர்கள். அவர்கள் எடுத்தவுடன் நீங்கள் தமிழில் பேசுகிறீர்கள், அவர்களுக்கு ஜெர்மானிய மொழியில் கேட்கிறது. அவர்கள் ஜெர்மானிய மொழியில் பதில் சொல்ல உங்களுக்குத் தமிழில் கேட்கிறது. இவ்வாறே பேசி வணிக வேலைகளை முடிக்க முடியுமென்றால் ஆங்கிலக் கல்வி தேவையற்றுப் போகாவிட்டாலும் அவசியம் குறைந்து விடுமல்லவா? அவரவர்கள் தங்கள் தாய் மொழியிலேயே பேசலாமே! இந்த நிகழ்வோட்டம் ஒரு காலத்தில் கேட்டால் அறிவியல் புனைகதை போல இருந்தது. ஆனால் இன்று [சந்தையில் வந்திருக்கும் பயணத்திற்கான மொழிபெயர்ப்புச் செயலிகளைப்](#) பார்த்தால் அந்த நாள் வெகு தூரத்திலில்லை என்றே தெரிகிறது.

english or 中文?

ஆங்கிலமா, சீன மொழியா?

அடுத்த குறிக்கோள் பத்தாயிரம் கன்பூசியஸ் சீன மொழி கல்விக் கூடங்கள் நிறுவுவதா?

[2008 இல் நிக்கோலஸ் ஆஸ்ட்லெர் ஃபோர்ப்ஸ் இதழில் ஒரு கட்டுரை எழுதினார்](#) இதில் சீன மொழியைக் கற்றுக்கொள்வதற்கு உலகெங்கிலும் 100 கன்பூசியஸ் கல்விக் கூடங்கள் நிறுவும் திட்டத்தின் மூன்றில் ஒரு பகுதியாக சீனா உள்ளது என்றார். அப்போது 23 நாடுகளில் இக்கல்விக் கூடங்கள் இருந்தன. இது 2010 ஆம் ஆண்டில் உலகம் முழுவதிலும் 100 மில்லியன் மக்கள் சீன மொழியைப் படிப்பதற்கான திட்டத்தின் ஒரு பகுதி. 2014 இல் [ஆறு கண்டங்களில் 480 க்கும் மேற்பட்ட கன்பூசியஸ் கல்விக் கூடங்கள் வந்து விட்டன](#). 2020 ஆம் ஆண்டுக்குள் 1,000 கன்பூசியஸ் கல்விக் கூடங்கள் நிறுவுவதே இவர்களது இப்போதைய குறிக்கோள்.

சீன மொழி உலகளாவிய மொழியாக ஆங்கிலத்தின் இடத்தைப் பிடிக்குமா?

[சிங்கப்பூரில் ஒரு பல்லாடக நிறுவனத்தை நடத்தும் லீ ஹான் ஷி](#), சீனாவில் வணிகம் செய்ய மேற்கத்திய வாடிக்கையாளர்களை அழைத்து வருகிறார். இவருக்கு வணிக ரீதியில் ஆங்கிலம் அவ்வளவு முக்கியமில்லை என்று கூறுகிறார். ஆங்கிலம் கற்றுக்கொள்ள வேண்டும் தான், ஆனால் நீங்கள் சீன மொழியையும் கற்றுக்கொள்ள வேண்டும். சீனாவின் பொருளாதார வலிமை வளர்ந்து வரும் நிலையில், சீன மொழி ஆங்கிலத்தை முந்தி விடும் என்று லீ நம்புகிறார். ஏற்கனவே இதற்கான அறிகுறிகள் உள்ளன என்றும் ஆங்கில மொழிச் சரிவு அமெரிக்க டாலரின் வீழ்ச்சியைத் தொடரும் என்றும் நம்புகிறார். நாடுகளிடையே பரிவர்த்தனைக்கான அடுத்த நாணயமாக டாலருக்குப் பதிலாக சீன நாணயமான ரென்மின்பி மாறினால், நீங்கள் சீன மொழியைக் கற்றுக் கொள்ள வேண்டும் என்று கூறுகிறார். பிரேசில்

மற்றும் சீனா போன்ற இடங்களில், அமெரிக்க டாலர் பயன்படுத்தாமல் ரென்மின்பியில் வணிகம் செய்து வருவதால், ஆங்கிலம் பயன்படுத்த வேண்டிய அவசியம் இல்லை.

மற்ற வெளிநாட்டு மொழிகளுடன் ஒப்பிடும்போது சீன மொழியைக் கற்றுக்கொள்வது மிகக் கடினம்தான். ஆனால் தொழில்நுட்பம் இதை எளிதாக்க உதவுகிறது. [டாக்டர் விக்டர் மேயர் தனது மொழி பற்றிய பதிவில்](#) நவீன எண்ணிம கருவிகள் சீன மொழியின் ஆய்வு மற்றும் பயன்முறையை எவ்வாறு மாற்றியமைக்கின்றன என்பதை விவாதிக்கிறார். எடுத்துக்காட்டாக, எழுதுவதை நம் சாதனங்கள் கையாளும் போது, நாம் ஏன் ஆயிரக்கணக்கான எழுத்துக்குறிகளை மீண்டும் மீண்டும் எழுதிப் பழக வேண்டும்? நம்முடைய நேரத்தை நடைமுறையில் பேசுவதிலும் நமது சொல்லகராதியைப் பெருக்குவதிலும் அதிகம் செலவழிக்கலாமே என்கிறார்.

6. தடை செய்யப்பட்ட கட்டலான் மொழி புத்துயிர் பெற்றது எப்படி?

ஸ்பெயின் நாட்டில் பாரசீலோனாவைச் சுற்றியுள்ள பகுதிகளில் கட்டலான் மொழி சிறுபான்மை மக்களால் பேசப்படுகிறது. வெற்றிபெற்ற ஆட்சியாளர்களால் பல நூற்றாண்டுகளாக ஒடுக்கப்பட்ட கட்டலான் மொழி இப்பொழுது 9 மில்லியன் மக்களால் பேசப்படுகிறது. உலகில் 150 க்கும் மேற்பட்ட பல்கலைக்கழகங்கள் இந்த மொழியை கற்றுத் தருகின்றன. 400 க்கும் மேற்பட்ட பத்திரிகைகள் இம்மொழியில் பிரசுரிக்கப்படுகின்றன.

பேரிடர்களை சந்தித்துப் பிழைத்து வந்த கட்டலான் மொழி

1714 ஆம் ஆண்டில் ஸ்பானிய துருப்புக்கள் பாரசீலோனாவை வெற்றி கண்ட பின், கட்டலோனியா அதன் தன்னாட்சி உரிமையை இழந்தது. கட்டலான் மொழி மீது கட்டுப்பாடுகளை விதித்து மத்திய அரசு ஸ்பானிஷ் மொழியை ஆட்சி மொழியாக ஆக்கியது.

உள்நாட்டுப் போரின் முடிவில் கட்டலோனிய மக்கள் ஒரு தனி உலகில் வாழ்ந்தார்கள். கட்டலான் மொழியைப் பயன்படுத்துவது தன்னுரிமையின் அடையாளமாகக் கருதப்பட்டது. வரலாறு பற்றி யாரும் பேசவில்லை. அரசியல் பற்றி யாரும் பேசவில்லை. ஆனால் கட்டலான் மொழியில் மக்கள் பேசிக்கொண்டிருந்தார்கள். இதை அவர்கள் அதிகாரபூர்வ ஸ்பெயின் நாடு மற்றும் ஆட்சியை எதிர்க்கும் அல்லது ஒதுங்கியிருக்கும் ஒரு அடிப்படை வழி என்று கருதினர். 19 ஆம் நூற்றாண்டில், கட்டலான் மொழி தேசியவாத கலாச்சார இயக்கத்தின் முயற்சியால் ஒரு இலக்கிய மொழியாக மறுபிறப்பு அடைய ஆரம்பித்தது. எனினும், இந்த மறுமலர்ச்சி நீடிக்கவில்லை.

1939 ஆம் ஆண்டில் பிரான்கோ சர்வாதிகார ஆட்சி வந்தவுடன் பொதுத்துறை நிர்வாகத்திலும், கல்வியிலும் கட்டலான் மொழி தடை செய்யப்பட்டது. பள்ளிகள், விளம்பரம், மத விழாக்கள் மற்றும் சாலைக் குறியீடுகளில் சிறுபான்மை மொழிகள் தடை செய்யப்பட்டன. புதிய அரசியல் நிலைமையை சுட்டிக்காட்டிய அதிகாரிகள், குறிப்பாகக் காவல் துறையினர், “பேரரசின் மொழியைப் பயன்படுத்துங்கள்” என்று மக்களைக் கட்டாயப் படுத்தினர். ஸ்பானிய மொழிடன் ஒப்பிடுகையில், கட்டலான் மொழி கௌரவத்தை இழந்தது. மேலும் சில மேல் வர்க்கத்தார் ஸ்பானிஷ் மொழியில் பேச ஆரம்பித்தனர். 1975 இல் ஸ்பெயினில் மீண்டும் மக்களாட்சி ஏற்பட்ட பின்னர், இது ஸ்பானிஷ் மொழியுடன் சேர்த்து அதிகாரபூர்வ மொழி, கல்விக்கான மொழி, மற்றும் ஊடகங்களின் மொழி என நிறுவப்பட்டது. இவை அனைத்தும் இதற்கு பெருமைக்குரிய பங்கை அளித்தன.

எண்ணிம காலத்தில் மொழித் தொழில்நுட்பங்கள்

நாம் எண்ணிம காலத்தின் ஆரம்ப கட்டத்தில் இருக்கிறோம். தொலைபேசி, இணையம், மின்னஞ்சல், திரைப்படங்கள் அல்லது இசையை எங்கிருந்தாலும் திறன்பேசி அல்லது கைக்கணினி மூலம் அணுகக்கூடிய வசதிகளை நாம் அனுபவித்து வருகிறோம். இந்த சாதனங்களின் பயன்பாடு இயற்கையாகி வருகிறது. முதலில் வந்த வணிகப் செயலிகளை விசைப்பலகைகள் மூலம் ஓட்ட வேண்டியிருந்தது. மொழித் தொழில்நுட்பத்தில் முன்னேற்றங்கள் வந்த பின்னர் வாய்ப் பேச்சாலேயே அணுக இயல்கிறது.

வெகு விரைவில் எந்த நேரத்திலும் எங்கிருந்தும் நமக்குத் தானியங்கி மொழிபெயர்ப்பு, உரையாடல்கள் மற்றும் ஆவணங்களின் சுருக்கம், தானியங்கியாக பேச்சை உரையாக்குவது, நிழல் படங்களுக்கு துணைத்தலைப்பு போன்ற சிக்கலான செயலிகளை நம்மால் அணுக முடியும். இந்தச் செயலிகள் அனைத்துக்கும் வெவ்வேறு மொழித் தொழில்நுட்பங்கள் (பேச்சை உரையாக்குதல், உரை ஒலி மாற்றி, மொழிபெயர்ப்பிகள், பாகுபடுத்திகள், சொற்பொருள் பகுப்பாய்வுகள் முதலியன) மற்றும் பொருத்தமான மொழி வளங்கள் (பேச்சு மற்றும் உரைத் தொகுதிகள், இலக்கணங்கள் முதலியன) ஒருங்கிணைந்து வேலை செய்ய வேண்டும்.

நவீன மொழித் தொழில்நுட்பங்களும் மொழியியல் ஆராய்ச்சிகளும் மொழியின் எல்லைகளை இணைக்கும் ஒரு பெரிய பங்களிப்பை செய்ய முடியும். திறன் பேசி போன்ற சாதனங்கள் மற்றும் செயலிகளுடன் இணைந்து, மொழித் தொழில்நுட்பம் எதிர்காலத்தில் ஒரு பொதுவான மொழியில் பேசாவிட்டாலும் ஒருவருக்கொருவர் எளிமையாக பேசுவதற்கும் வியாபாரம் செய்வதற்கும் உதவ முடியும்.

மொழித் தொழில்நுட்பத் தீர்வுகள் இறுதியில் வெவ்வேறு மொழிகளுக்கு இடையில் ஒரு தனிப்பட்ட பாலமாக செயல்படும். மேலும், தகவல் மற்றும் புதிய செயலிகளுக்கான உலகளாவிய அணுகல் குறிப்பிட்ட மொழியின் பயன்பாட்டால் வரையறுக்கப்படாது. இருப்பினும், சந்தையில் தற்போது கிடைக்கும் மொழித் தொழில்நுட்பங்கள் மற்றும் பேச்சு செயலாக்கக் கருவிகள் இன்னும் இந்த லட்சிய இலக்கை அடையவில்லை. ஐரோப்பிய ஒற்றுமைக்கு மொழித் தொழில்நுட்பத்தின் ஆழமான தொடர்பையும் முக்கியத்தையும் 1970 களின் பிற்பகுதியிலேயே ஐரோப்பிய ஒன்றியம் உணர்ந்தது. அதன் முதல் ஆராய்ச்சி திட்டங்களுக்கு நிதியுதவி செய்தது.



ஸ்பெயின் நாட்டில் சிறுபான்மை மொழி பேசும் வட்டாரங்கள்

கட்டலான் மொழியின் இயல் மொழித் தொழில்நுட்பம்

கட்டலான் மொழியின் இயல் மொழித் தொழில்நுட்பம், மக்கள் தொகை மற்றும் மொழியின் அரசியல் செல்வாக்கு ஆகியவற்றைக் கருத்தில் கொண்டால், எதிர்பார்ப்பதை விடச் சிறந்த வடிவில் உள்ளது. மற்ற வளங்களுக்கிடையில், கட்டலான் மொழியில் ஒரு உயர்தர கைமுறையாகக் குறியீடு செய்த 55 மில்லியன் சொற்தொகுப்பு (manually annotated corpus) உள்ளது. இது பத்தாண்டுகளுக்கும் மேலாக பார்சிலோனா பல்கலைக்கழகம், கட்டலான் மொழியறிவியல் மற்றும் பொது மொழியியல் துறை முனைவர் வக்சீம் ரபேல் முன்யோசனையுடன் தொகுத்தது. ஸ்பானிஷ் மற்றும் கட்டலான் மொழிகளில் இரண்டு பதிப்புகளாக வெளியிடப்படும் தினசரி பத்திரிகையின் வெளியீடும் மொழிபெயர்ப்பு இணை உரைகளுக்கு ஒரு முக்கிய ஆதாரமாக உள்ளது.

மொழித் தொழில்நுட்பத்தின் பலதுறை இயல்பு காரணமாக, தொழில்நுட்ப உலகில் (கணினி அறிவியல் மற்றும் பொறியியல் துறைகள்) மற்றும் மனிதநேயத் துறையில் (மொழி மற்றும் மொழிபெயர்ப்பியல் துறைகள்) இருந்து ஆராய்ச்சிக் குழுக்கள் வந்து சேருகின்றன. குழுக்களில் பெரும்பாலானவை பல்கலைக்கழகங்களைச் சேர்ந்தவை. அடிப்படை அல்லது செயல்முறை சார்ந்த ஆராய்ச்சியில் ஈடுபடுகின்றன. தொழில்நுட்பக் கைமாற்றத்தில் இதில் ஒரு குழு தேர்ச்சி பெற்றிருக்கிறது, ஆனால் எல்லாக் குழுக்களும் இதை ஏதாவது ஒரு விதத்தில் சமாளிக்கின்றன.

உள்ளூர் அரசாங்கம், மாநில அரசு அல்லது ஐரோப்பிய பல்கலைக்கழகம் (Universidad Europea) ஆகியவற்றிலிருந்து திட்டங்களுக்கு நிதி வருகிறது. மற்ற நிதி ஆதாரங்கள் நிறுவனங்களுடன் அல்லது நிர்வாகத்துடன் ஒப்பந்தங்களாகும். கட்டலானுக்கான பெரும்பாலான ஆராய்ச்சிகள் எழுதப்பட்ட உரையைப் பற்றித்தான். பதினான்கு குழுக்களில் பத்து உரைத் தரவுகளில் கவனம் செலுத்துகின்றன. மூன்று குழுக்கள் உரையிலும் பேச்சிலும் வேலை செய்கின்றன. ஒரு குழு பேச்சில் மட்டுமே வேலை செய்கிறது. பேச்சில் அடிப்படைப் பகுதிகள் பேச்சுத் தயாரிப்பு (speech synthesis) மற்றும் பகுப்பாய்வு ஆகும்.

இந்தக் கூட்டு முயற்சியின் காரணமாக, மொழி செயலாக்கத்திற்கான அடிப்படைக் கருவிகள் மற்றும் ஆதாரங்கள் ஆகியவை கட்டலான் மொழிக்கு உள்ளன. இருப்பினும் அவை எப்போதும் பரவலாகத் தெரியவருவதில்லை. சில சமயங்களில் ஆராய்ச்சி சமூகம், தொழில் நிறுவனங்கள் அல்லது பொதுமக்களுக்குக் கிடைப்பதில்லை. பல்வேறு குழுக்களிடையே (ஆராய்ச்சிக் குழுக்கள், அரசாங்கம் மற்றும் நிறுவனங்கள்) நல்ல ஒருங்கிணைப்பு மற்றும் ஒத்துழைப்பு எந்தவொரு முயற்சிக்கும் தேவைதான். இருப்பினும் கட்டலான் போன்ற ஒரு சிறிய மொழிக்கான தொழில்நுட்பம் உருவாக்குவதில் இது மிக அவசியம். மொழி வளங்கள் விலை உயர்ந்தவை, நிதியோ குறைவு, மற்றும் சமூகம் முயற்சிகளை திரும்பச் செய்யவும் சிதற அடிக்கவும் இடம் கிடையாது.

7. “ஆளும் மொழியே வாழும்; மற்றவை மாளும்” இதுதான் நியதியா?

ஆக்ஸ்போர்ட் மொழியியல் பேராசிரியர் ஜீன் அட்சிசன் சொல்கிறார், “ஒரு மொழியின் பரவல் அதைப் பயன்படுத்துபவர்களுடைய சக்தியைச் சார்ந்தது, அம்மொழியின் உள் அம்சங்களைப் பொருத்தது அல்ல.” தமிழின் பயன்பாட்டைக் கட்டாயமாக்கும் ஒரு நாட்டைத் தமிழர்கள் அமைத்தால் மட்டுமே தமிழ் வளரவும் செழிக்கவும் முடியும் என்று சிலர் உறுதியாக நம்புகின்றனர். எடுத்துக்காட்டாக இதோ ஒரு கட்டுரை. [இராஜேஸ்வரி பாலசுப்பிரமணியம் எழுதிய தமிழ் மொழியின் எதிர்காலம்.](#)

“தமிழ் மொழி, எதிர்காலத்தில் வாழ்வது நிலப்பரப்பிலோ, மக்கள் தொகையிலோ அல்ல; மாறாக அதன் ஆளும் வல்லமையில் மட்டுமே தங்கியுள்ளது. ஆளும் மொழியே வாழும்; மற்றவை மாளும்.”

ஸ்பானிய, ஆங்கிலேயர்கள் மற்றும் போர்த்துகீசியர்கள் குடியேற்றத்தின் மூலம் தங்கள் மொழிகளை ஆதிக்கம் செலுத்தும் உலகளாவிய மொழிகளாக ஆக்கியுள்ளதை வைத்து இவ்வாறு நினைப்பதாகத் தோன்றுகிறது. இன்னும் சிலர் ஹீப்ரு மொழியை மேற்கோள் காட்டுகின்றனர். இது தினசரி பேச்சு வழக்கில் இல்லாமல் யூத வழிபாட்டு மொழியாகியிருந்தது. பின்னர், 19 ஆம் நூற்றாண்டில், நவீன ஹீப்ருவாக மறுவாழ்வு பெற்று இன்று இஸ்ரேல் நாட்டின் ஆட்சி மொழியாகவும் இலக்கிய மொழியாகவும் புத்துயிர் பெற்றது. யூதர்களுக்கு இஸ்ரேல் என்ற தனி நாடு அமைத்ததனால் மட்டுமே நடந்தது. இல்லையெனில் நடந்திருக்கவே முடியாது என்பதென்னவோ உண்மைதான்.

ஆனால் நாம் இதன் வேறொரு கோணத்தையும் பார்க்க வேண்டும். நார்டிக் நாடுகள் என்பது ஸ்வீடன், நார்வே, பின்லாந்து, ஐஸ்லாந்து, டென்மார்க், கிரீன்லாந்து சேர்ந்தது. இந்த நாடுகளுக்கு தேசிய மொழிகள் இருந்த போதிலும், [அறிவியல், உயர்கல்வி மற்றும் வணிக மொழியாக ஆங்கிலத்தை ஏற்றுக்கொள்வதற்கு அழுத்தம் அதிகரித்துள்ளது.](#) நார்டிக் அறிஞர்கள் தங்கள் ஆய்வு முடிவுகளை பல்நாட்டு பத்திரிகைகளில் வெளியிட வேண்டிய கட்டாயத்தில் உள்ளனர். இதுபோலவே, அயர்லாந்து என்ற தனி நாடு அமைத்து ஐரிஷ் மொழியை அதன் ஆட்சி மொழியாக ஆக்கிய பின்னும் ஐரிஷ் மொழி பேசுவோர் மற்றும் கற்போர் குறைந்து வருவதைத் தடுக்க இயலவில்லை.

வணிக ரீதியாக முக்கியமான மொழிகளிலேயே பெரிய நிறுவனங்கள் கவனம் செலுத்துகின்றன

ஆங்கிலம் அதன் தனி பகுப்பில் உள்ளது (என்னுடைய மருமகளின் தமிழ்கில மொழியில் சொல்லப்போனால் “அது வேற லெவல்”). உதாரணமாக, [இந்த உரைத்தொகுப்பு மொழியியல் வலைப்பதிவு பாருங்கள்.](#)

“இயல்மொழி ஆய்வு பற்றிய திறந்த இரகசியம் இதுதான். அது மிகவும் ஆங்கிலத்தை மையமாகக் கொண்டுள்ளது. மொழியியலாளர்கள் மிக அதிகமாக வேலை செய்துள்ள மொழியாகவும், கணினி அறிவியல் திட்டங்களுக்கான மிகச் சிறந்த வளங்களைக் கொண்டிருக்கும் மொழியாகவும் ஆங்கில மொழி உள்ளது. கணினி அறிவியலில் எப்போதுமே இன்னமும் தரவுகள் இருந்தால் இன்னமும் நல்லது.”

[இந்த ஆய்வறிக்கையில் விவரிக்கப்பட்டுள்ளபடி](#), பிற வணிகரீதியாக முக்கியமான மொழிகளான பிரஞ்சு மற்றும் ஜெர்மன் போன்ற மொழிகளுக்கு கூகிள், முகநூல், அமேசான், மைக்ரோசாப்ட் போன்ற பெரிய நிறுவனங்களிடமிருந்து நல்ல ஆதரவு கிடைக்கும். ஒரு பெரிய சாத்தியமான செல்வந்த சந்தை இருந்தால் பெரிய சந்தை நிறுவனங்கள் அந்த மொழிகளைக் கவனித்துக்கொள்வார்கள். ஆனால் சந்தை அதைச் செய்யாவிட்டால் மற்ற மொழிகளுக்கு முறையான தொழில்நுட்பங்களை யார் வழங்குவார்? சிறிய மொழிகளில் சிறிய நாடு, மாநிலம் அல்லது வட்டாரப் பொருளாதாரங்கள்தான் உள்ளன. எனவே தொழில்நுட்பத்திற்கான தனியார் அல்லது பொது நிதிகளுக்கான வாய்ப்புகள் குறைவாகவே உள்ளன. இயல்மொழி ஆய்வைப் பொறுத்தவரை அனைத்து மொழிகளும் கடினமான மொழிகள்தான். ஆனால் அவைகளின் நிதி நிலைமைகள் மலையிலிருந்து மடு போன்று வேறுபடுகின்றன.

இன்னும் பத்து ஆண்டுகளில் அயர்லாந்திலேயே ஐரிஷ் முக்கிய மொழியாக இருக்காது

பல நூற்றாண்டுகளாக அயர்லாந்தில் பேசப்படும் மொழி ஐரிஷ் மட்டுமே. 17 ஆம் நூற்றாண்டில் இங்கிலாந்தின் அரசர் முதலாம் ஜேம்ஸ் நிலத்தைக் கைப்பற்றுவதற்காக பலரை அனுப்பி, கலகங்களை அடக்கி, நாட்டைக் குடியேற்றமாக்கினார். ஆங்கில மேலாதிக்கத்தின் கீழ் ஏற்கனவே மொழி மறைந்து கொண்டிருக்க 1800 களின் பஞ்சம் நாட்டின் ஐரிஷ் பேசுபவர்களை வெளியேற்றியது. தனி நாடாகப் பிரிந்த பிறகு, அயர்லாந்தில் ஐரிஷ் ஆட்சி மொழியாக ஆயிற்று.

2006 மற்றும் 2011 கணக்கெடுப்புகளின் தரவின் அடிப்படையில் சமூகத்தின் மொழிப் பயன்பாட்டை இந்த ஆய்வு ஒப்பிடுகிறது. ஐரிஷ் பேசும் பகுதிகளில் மொத்தம் 155 வட்டாரங்களில் 21 இல் மட்டுமே மக்கள்தொகையில் மூன்றில் இரு பகுதியினர் ஐரிஷ் பேசுகின்றனர். இது முந்தைய கணக்கெடுப்பில் 24 ஆக இருந்தது பின்னர் குறைந்து விட்டது. அயர்லாந்தில் ஐரிஷ் மொழி பேசும் வட்டாரங்களில் மொழியின் தேய்மானம் 2007 ஆய்வில் எச்சரித்ததை விட மிக வேகமாக நடைபெறுகிறது. பத்து ஆண்டுகளில் ஐரிஷ் மொழி பேசும் பகுதிகளில் எங்குமே அது முக்கிய மொழியாக இருக்கும் என்று வாதிடுவது கடினம்.

“ஐரிஷ் மூலம் தங்கள் குழந்தைகளை வளர்க்கத் தீர்மானிக்கும் தாய்மார்களும் தந்தையர்களும் தான் உண்மையாக மொழியைத் திட்டமிடுபவர்கள்” என்று அவர் கூறினார்.



அயர்லாந்தும் வட அயர்லாந்தும்

ஆனால் வட அயர்லாந்திலோ அரசியல்வாதிகளைவிட மக்கள் ஒரு தப்படி முன்னால் இருக்கிறார்கள்

ஐரிஷ் மொழி இங்கிலாந்தின் பகுதியான வட அயர்லாந்தில் அனேகமாக மறைந்தே விட்டது. இருப்பினும், மாகாணத்தின் பல சமூகங்களில், சில குடும்பங்கள் மொழியைக் காப்பாற்ற ஒன்று சேர்ந்தனர். இந்த முயற்சி ஒரு கிராமப்புறப் பகுதியில் தொடங்கியது. 1992 இல் ஒரு பத்துப் பன்னிரண்டு குடும்பங்கள் தங்கள் சேமிப்பைச் சேர்த்து ஒரு ஐரிஷ் அறிமுக நிலைப் பள்ளியை அமைத்தனர். மாகாணத்தில் இது முதல் பள்ளியாகும்.

இந்தப் பள்ளியைத் தொடங்குவதற்கு உதவியவர்களுள் ஒருவர் நினைவுகூர்ந்தார், “ஒரு சிலப் பெற்றோர்களுக்குத்தான் இந்தத் தைரியம் இருந்தது. அரசு நிதி எதுவும் கிடையாது. பிரிவினையிலிருந்து ஐரிஷ் மொழியானது நாட்டு மொழியாகக் கருதப்படுவதில்லை. பழைய அணுகுமுறை ஐரிஷ் மொழியை ஒரு வெளிநாட்டு மொழியாகக் கருதி ஒவ்வொரு மட்டத்திலும் ஊக்கமளிப்பதில்லை.”

இந்தக் குடும்பங்களோ விடாப்பிடியாகத் தொடர்ந்து முயற்சி செய்தன. இன்று இந்த காரன்டோகர் சமுதாய சங்கம் ஒரு வளர்ந்து வரும் ஐரிஷ் மொழிப் பள்ளியை நடத்தி வருகிறது.

“மொழியில் பெரும் வளர்ச்சியை நாங்கள் கண்டிருக்கிறோம்,” என்று மையத்தில் ஒரு மேலாளரான லியோன் கூறினார். “ஐரிஷ் மொழி எவருக்கும் எதிரானதல்ல. எங்கள் மொழியைப் பேசுவதற்கு எங்களுக்கு உரிமை உண்டு. வேறு கருத்து யார் சொன்னாலும் அது மிகவும் குறுகிய மனப்பான்மை என நான் கருதுகிறேன்.” இந்தக் கல்வி இயக்கம் மற்ற இடங்களிலும் பரவியது. இறுதியில் வட அயர்லாந்து அரசாங்கம் ஒருசில பள்ளிகளில் ஐரிஷ்-மொழி கற்கும் திட்டங்களுக்கு நிதி வழங்க ஒப்புக்கொண்டது. இன்று, முதல் நிலை மற்றும் இரண்டாம் நிலையில் 39 ஐரிஷ் மொழிப் பள்ளிகள் உள்ளன.

“அரசியல்வாதிகளைவிட மக்கள் இங்கே ஒரு தப்படி முன்னால் இருக்கிறார்கள் என்று நான் நினைக்கிறேன்.”

சுருக்கமாகச் சொல்வதானால் “உங்கள் மொழியை மற்ற நாடுகள் இனங்களின் மீது கட்டாயப் படுத்தித் திணிக்க வேண்டுமென்றால் அது ஆளும் மொழியாக இருப்பது அவசியம். ஆனால் உங்கள் சமூகத்திலேயே ஆழமாக வேருன்ற வேண்டுமென்றால் அதற்கு சமூக முனைப்பும் தகவல் தொழில்நுட்பமும் தான் உதவ முடியும்.”

8. புதிய தலைமுறையின் மரபணுவே எண்ணிமத்தால் ஆனது போலுள்ளது

நுகர்பொருள் ஆய்வக அறிக்கையின்படி இந்தியாவில் 18 வயதிற்கு உட்பட்ட இளையவர்கள் சுமார் 200 மில்லியன் உள்ளனர், அவர்களில் 69 மில்லியன் பேர் நுகர்ப்புறங்களில் வாழ்கின்றனர். 1981 முதல் 1995 வரை பிறந்த தலைமுறையை ஆங்கிலத்தில் மில்லேனியல் என்று சொல்கிறார்கள். இவர்களுக்கு தகவல்தொடர்பு, ஊடகம், எண்ணிம தொழில்நுட்பங்கள் ஆகியவற்றுடன் நல்ல பரிச்சயம் உண்டு. இவர்களுக்கு அடுத்து வந்த 1996 முதல் 2010 வரை பிறந்தவர்களை இணைய அல்லது எண்ணிம தலைமுறை என்றே சொல்லலாம். இந்த இளைய தலைமுறையின் குழந்தைப் பருவம் அவர்களுடைய பெற்றோர்களுடையதை விட மிகவும் வித்தியாசமானது.

தொழில்நுட்பம் இளம் இந்தியாவை முற்றிலும் மாற்றிவிட்டது. முன்னால் ஆடம்பரமாகக் கருதப்பட்ட சாதனங்கள் வெகு விரைவில், கல்வியிலும் சொந்த வாழ்க்கையிலும் மிகுந்த தாக்கத்தை ஏற்படுத்தும், இன்றியமையாதவையாக ஆகிவிட்டன. டாடா நிறுவனத்தின் இணைய தலைமுறைக் கருத்தாய்வில் 15 நகரங்களில் சுமார் 12,000 மாணவர்கள் கேள்விகளுக்கு விடைகொடுத்தார்கள். திறன்பேசி (75%) மிகவும் பரவலாகப் பயன்படுத்தப்படும் சாதனம். அடுத்து மடிக்கணினிகள் (67%) மிகவும் பின்னால் இல்லை. கைக்கணினிகள், மின்னூல் படிப்பிகள், விளையாட்டு முனையங்கள், திறன் கைக்கடிகாரங்கள் மற்றும் மெய்நிகர் உண்மை தலைக்கவசம் ஆகியவை பெருநகரங்களிலும் மற்ற நகரங்களிலும் அந்த வரிசையில் பின்பற்றப்படுகின்றன. பெண்களுக்கு மின்னூல் படிப்பிகள் முன்னால் உள்ளன, ஆண்களுக்கு விளையாட்டு முனையங்கள் முன்னால் உள்ளன. பதிலளித்தவர்களில் 26% தினமும் குறைந்தது ஒரு மணிநேரத்தை இணையத்தில் செலவிடுகின்றனர். 27% இளைஞர்கள் தங்களுக்கு வந்த தகவல்களுக்கு 5 நிமிடங்களுக்குள் பதிலளிப்பதாகச் சொல்கிறார்கள்.

15 வயதுக்குக் கீழே உள்ள தலைமுறையைப் பற்றிக் கேட்கவே வேண்டாம்!

நுகர்ப்புறங்களில் 15 வயதுக்குக் கீழே உள்ள 30 மில்லியனுக்கும் அதிகமானோரிடம் கைப்பேசி உண்டு. மேலும் 11 மில்லியன் பேர் குடும்ப உறுப்பினர்களின் கைப்பேசியைப் பகிர்ந்து கொள்கின்றனர். இந்தத் தலைமுறைக்கு திறன்பேசி இல்லாமல் வாழ்க்கை என்ன என்றே தெரியாது. இவர்கள் எப்போதுமே இணையத்திலுள்ளனர். தங்களுக்கு எம்மாதிரி எதிர்காலம் வேண்டுமென்றும், அந்த எதிர்காலத்தை உருவாக்க எந்த வழியில் செல்ல வேண்டுமென்றும் தெரியுமென்று இவர்கள் நினைக்கிறார்கள். இதுதான் திறன்பேசியுடனே வளர்ந்த முதல் தலைமுறையாகும். இணையம் இவர்களது அடிப்படைத் தேவைகளில் ஒன்றாகும். இவர்கள் இணையத்துடன்தான் தூங்கச் செல்கிறார்கள், இணையத்துடன்தான் எழுந்திரிக்கிறார்கள். மில்லேனியல்களுக்கோ அவர்கள் குழந்தைப் பருவத்தில் கைப்பேசிகள் இல்லை. தேவைப்படும்போது குடும்பத்தினருடன் பகிர்ந்து கொண்ட மேசைக் கணினியில் இணையத்துக்குச் செல்ல வேண்டியிருந்தது. எண்ணிம தலைமுறையோ ஏதாவது ஒன்றைக் கற்றுக்கொள்ள அல்லது தேட விரும்பினால், உடனடியாக இணையத்தைப் பயன்படுத்துகிறார்கள். இவர்கள் முக்கியமாக நிழல்படங்களில் இருந்துதான் தகவலைச் சேகரிக்கிறார்கள். கூகிளை விட யூடியூபை ஒரு தேடுபொறியாகப் பயன்படுத்துகிறார்கள். யூடியூபில் உள்ள பல்லாயிரக் கணக்கான நிகழ்படப் பயிற்சிகளைப் பாருங்கள்.

இவர்களுக்கு விதிமுறைகளைக் கடைப்பிடிக்கும் பொறுமை கிடையாது. ஒரு புள்ளியிலிருந்து மற்றொரு புள்ளிக்கு உள்ள நேரடிப் பாதையையே எல்லா நேரங்களிலும் விரும்புகின்றனர். அமெரிக்காவில் பகுதிநேர வேலை தேடுவோர் என்னென்ன வேலைகள் செய்ய வேண்டுமென்று எழுதியிருப்பதை ஒரு முழு நிமிடம் படிக்கக்கூடப் பொறுமை கிடையாது. மேலும் 97% பேர் விண்ணப்பத்தை முழுவதும் பூர்த்தி செய்வதில்லை.



எண்ணிம தலைமுறை

இளைய தலைமுறைகளைப் பற்றி நல்ல விதமாகச் சொல்ல எதுவுமே இல்லையா?

உண்டு! சமீபத்தில் இங்கிலாந்தின் வர்கி அறக்கட்டளை பல்வேறு நாடுகளிலிருந்து இளைய தலைமுறையினர் 20,000 பேர்களுடம் கேள்விகளைக் கேட்டனர். இவற்றில் ஆயிரம் பேர் இந்தியர்கள். உலகளாவிய மட்டத்தில் இந்தத் தலைமுறையில், தன்னை மற்றும் தன் குடும்பம் மற்றும் நண்பர்களை மட்டுமே கவனிக்காமல், சமுதாயத்தில் ஒரு பரந்த பங்களிப்பை செய்வது மூன்றில் இரண்டு மடங்கானவர்களுக்கு (67%) முக்கியம் என்று இந்த ஆய்வு கூறுகிறது. பள்ளி, கல்லூரி என்ற முறைப்படியான கற்றல் சூழலில் இருந்து விடுபட்டு இணைய கற்றலில் ஈடுபாடு கொள்வது இந்தியாவில் பதினம் வயது இளைஞர்களுக்கு காணப்படும் ஒரு சுவாரசியமான போக்கு. ஆகவே தனக்குத்தானே மற்றும் கூடிக் கற்றல் இவர்களால் பரவலாகப் பயன்படுத்தப்படுகிறது. எனவே இளைய தலைமுறையினருக்கு எடுத்துச் செல்ல வேண்டும் என்றால் மொழியை எல்லா எண்ணிம ஊடகங்களிலும், எல்லா சாதனங்களிலும் கிடைக்கச் செய்ய வேண்டும்.

நடைமுறை எடுத்துக்காட்டு – ஒடியா விக்கிபீடியா

ஒடியா விக்கிபீடியா எப்படி அந்த மொழியின் இணையப் போக்கை மாற்றியது என்று சுபாஷிஷ் பாணிக்கிராகி விவரிக்கிறார். உங்கள் மொழி எவ்வளவு பழமையானது என்பது முக்கியமல்ல, ஆனால் எல்லா ஊடகங்களிலும் எவ்வளவு நன்றாகப் பயன்படுத்தப்படுகின்றது என்பது மட்டும்தான். அவற்றின் புகழ்பெற்ற கடந்த காலத்தையும் மற்றும் இலக்கியப் பாரம்பரியத்தையும் காரணமாகக் கொண்டு மொழிகள் வளர்வதில்லை. இந்த மரபு நிச்சயமாக ஒடிய மொழிக்கு எந்த ஒரு பலனையும் அளிக்கவில்லை. மக்களுக்கு மொழியை எடுத்துச் சென்ற எண்ணிமக் கருவிகள்தான் உதவின.

ஆனால் இதுமட்டும் போதாது. மெய்நிகர் மற்றும் மிகைப்படுத்தப்பட்ட மெய்ம்மை, இயத்திரக் கற்றல், ஆழமான கற்றல் மற்றும் செயற்கை நுண்ணறிவு ஆகியவையே மூச்சுக் காற்றாக இருக்கும் இந்த உலகில், அச்ச வெளியீட்டை மட்டுமே நம்பி மொழிகளை வளர்க்க முடியும் என்று கனவு காண்பது பைத்தியக் காரத்தனம் இல்லையா? குழுத் தொகுப்பில் இணையத்தில் வெளியிடும் போது அந்த உள்ளடக்கத்தின் நம்பகத்தன்மையில் சிறிய ஐயப்பாடு உள்ளதுதான். ஆனால் ஒரு சமூகத்தின் அறிவை ஆவணப்படுத்துவதற்கு மக்களின் ஒரு பெரிய பிரிவை ஈடுபடுத்த மிக விரைவான, எளிதான மற்றும் சம உரிமை வழி இதுதான் என்பதை நீங்கள் புரிந்து கொள்ள வேண்டும். விக்கிபீடியா பல மொழிகளுக்கு புத்துயிர் கொடுப்பதில் வெற்றி பெற்றதன் முக்கிய காரணம் அதன் தேர்ந்தெடுத்த உள்ளடக்கத்தை கூட்டுசேகரத்தில் உருவாக்கும் மாதிரிதான் (curated content-crowdsourcing model). சிலசமயங்களில், சிறிய மொழிகளுக்கு வழக்கமான அச்ச மற்றும் மின்னணு ஊடகத்தை உருவாக்க மிகவும் செலவு செய்ய வேண்டும். இலவச செயலிகள் மற்றும் சமூக ஊடகங்கள் மூலம் ஒரு பைசா கூட செலவழிக்காமல் மில்லியன் கணக்கானவர்களுக்கு உள்ளடக்கத்தை எடுத்துச் செல்ல இது ஒரு சிறந்த கருவியாகும்.

தாய்மொழியைப்பற்றி அடுத்த தலைமுறையின் மனப்பாங்கு என்ன?

உங்களுக்கு புதிய தலைமுறையின் தாய்மொழி பற்றிய மனப்பாங்கு ஓரளவு தெளிவாகத் தெரிய வேண்டுமா? இந்த யூடியூப் நிழல்படத்தைப் பாருங்கள். இதை 5 மில்லியனுக்கு மேல் பார்த்துள்ளார்கள். இதைவிட முக்கியமாக இதன் கிழே உள்ள கருத்துரைகளைப் படியுங்கள்.

- என் பாட்டி என் அப்பாவிடம் பேசும் போது அவர்கள் என்ன பேசிக்கொள்கிறார்கள் என்று எனக்குப் புரியாது என்பதை என்னால் எடுத்துக்கொள்ள முடியவில்லை.
- தாய் மொழியில் பேசுவதற்கான திறமை வீட்டிலேயே தொடங்குகிறது. நாங்கள் வளரும் பருவத்தில் வீட்டில் தாய் மொழியில் பேசாவிட்டால் அப்பா எங்களிடம் பேச மறுத்து விடுவார். அதற்கு நான் இப்போது மிகவும் நன்றியாக உணர்கிறேன்.
- அமெரிக்கா பல கலாச்சாரங்களின் கலப்பு என்பதால் உங்கள் தாய் மொழியையும் பேசி வளர்ந்த உங்களுக்கு வாழ்க்கையில், உங்கள் சக அமெரிக்கர்களுக்கு இல்லாத, சில நன்மைகள் உண்டு.
- நான் தாய் மொழியைக் கற்றுக்கொள்ள மறுத்தால் என் பாட்டிக்குக் கடும் கோபம் வரும். ஒவ்வொரு இரவும் அவள் என்னை பயிற்சி செய்யக் கட்டாயப் படுத்தும் போது எனக்குக் கடும் கோபம் வரும். ஆனால் நான் என்னுடைய தாய்மொழியைப் பேச முடிகிறது என்று இப்போது மகிழ்ச்சியடைகிறேன்.
- நான் என் தாய் மொழியைக் கற்றுக்கொள்வதற்கு முயற்சி செய்வதனால் என் குடும்பத்தைப் பெருமைப்படுத்துகிறேன். இந்த அருமையான மொழியை என் குழந்தைகளுக்கும் கொண்டு சேர்க்க வேண்டும்.

புதிய தலைமுறைக்கு தமிழில் ஆர்வம் வளர்ப்பது எப்படி?

புதிய தலைமுறையை எண்ணிம ஊடகங்கள் மூலம்தான் அணுக இயலும். அமெரிக்காவில் கடைகளில் சாளரத்தில் “உதவி தேவை” என்று எழுதி வைப்பார்கள். இவர்கள் அதைக் கவனிக்க மாட்டார்களாம். வரி விளம்பரங்களையும் படிப்பதில்லையாம். இந்தத் திரைப் பழக்க அடிமைகளின் கவனத்தை ஈர்க்க, நீங்கள் சம்மந்தப்பட்ட ஒவ்வொரு இணைய தளத்திலும் இடுகையிடவும் என்று ஆலோசனை கூறுகிறார்கள்.

இரண்டாவது, இவர்களுக்கு இணையத்தில் பயிற்சி நிழற்படங்கள் கொடுத்தால், தானே கற்றுக் கொள்வார்கள் மற்றும் பங்களிப்பார்கள். இந்த அடிப்படையிலுள்ள குழுவிலிருந்துதான் அடுத்த தலைமுறைக்கான எண்ணங்கள், ஈடுபாடுகள் மற்றும் தலைவர்கள் உருவாக வேண்டும். ஆகவே இந்த இணைய தலைமுறைக்கு தமிழ் மொழியைக் கொண்டு செல்வதும் ஆர்வம் வளர்ப்பதும் எப்படி என்பதுதான் இப்போதைய சவால்.

9. உங்கள் பிள்ளைகளை இயந்திர மனிதர்களாக வளர்க்கிறீர்களா?

தாய்மொழியை இழந்தால் தாயை இழந்ததுபோல் பரிதவிப்போம் என்பது மிகையாகாது

“பல புலம்பெயர்ந்த தமிழ்ப் பிள்ளைகளைப் போலவே வளரும் காலத்தில் நான் தமிழ் பேசவில்லை. என் கல்லூரிப் பருவத்திலும் வயதுவந்த பின்னும் என் பெற்றோரைத் திட்டிக் கொண்டிருந்தேன், ஏன் எனக்குக் குழந்தைப் பருவத்திலேயே தமிழ் சொல்லிக் கொடுக்கவில்லை என்று. வயது வந்தபின் ஒரு மொழியைக் கற்றுக்கொள்வது மிகவும் கடினமாக இருந்தது. அதே நேரத்தில் என் வாழ்க்கைக்கு நான்தானே பொறுப்பு. நான் தமிழ் பேச விரும்பினேன்.

தாய்மொழியைப் பேச வேண்டுமென்ற என் உணர்ச்சிகளில் ஒரு இக்கட்டான கட்டத்தை நான் அடைந்து விட்டேன். நான் எப்படியாவது தமிழ் பேச வேண்டும். வெட்கம், சங்கடம், அவமானம் போன்ற உணர்ச்சிகளுக்கு என் வாழ்க்கையில் இந்தக் கட்டத்தில் இடமில்லை என்று நான் முடிவு செய்தேன். எங்கு வேண்டுமானாலும் யாரிடம் வேண்டுமானாலும் தமிழ் பேசுகிறேன். அடிக்கடி தவறு செய்கிறேன். அதைக் கேட்டுச் சிரிக்கிறார்கள். அவ்வாறு நடக்கும்போது, அவர்கள் சிரித்த அக்கணத்தில் நான் என்ன பேச முயற்சி செய்தேன் என்பதைக் கவனத்தில் கொள்வேன். பின்னர் என்ன தவறு செய்தேன் என்று என் அம்மாவிடம் கேட்பேன். அல்லது, அவர்கள் என் நண்பர்கள் என்றால், நேரடியாக அவர்களிடமே கேட்பேன்.

எனக்கு ஏற்கெனவே மூச்சு விட நேரம் கிடையாது. மற்ற வேலைகளை ஒதுக்கி வைத்து விட்டுத் தமிழ் கற்றுக்கொள்ள நேரம் உருவாக்க வேண்டியிருந்தது. ஒரு சமயத்தில் வாரம் மூன்று தமிழ் வகுப்புகள் போய்க்கொண்டிருந்தேன். என் அம்மாவுடன் ஒரு மணிநேரம் ஸ்கைப். ஒரு மணிநேரம் டொராண்டோவில் ஆசிரியருடன் ஸ்கைப். ஞாயிற்றுக்கிழமை உள்ளூர் தமிழ்ப் பள்ளியில் ஒரு மணிநேரம், 8 அல்லது 9 வயது சக மாணவர்களுடன்.

வயது வந்தபின் தமிழ் மொழியைக் கற்றுக் கொள்வது கடினமாக இருப்பதால் நான் நிறைய அழுகிறேன். உங்களுடைய பெற்றோர் என்னுடைய பெற்றோர் போல் என்றால், நீங்கள் உங்களுக்கே மிக உயர்ந்த எதிர்பார்ப்புகளை ஏற்படுத்திக்கொள்ளுமாறுதான் வளர்க்கப்பட்டிருப்பீர்கள். பெரும்பாலும் அநியாயமான அதிக உயர் எதிர்பார்ப்பு. அந்த எதிர்பார்ப்புகளை அடைய இயலவில்லை என்றால் சில நேரங்களில் உலகமே முடிவது போலிருக்கும். தாய்நாட்டுடனும், மக்களுடனும் இருக்கும் தொடர்பைக் காப்பாற்றிக் கொள்ளவேண்டுமென்ற என் ஆழ்ந்த வேருன்றிய ஆசையுடன், தமிழ் மட்டுமே பேசும் எனது குடும்பத்தாருடனும் சரளமாக உரையாட விரும்பும் உணர்ச்சிபூர்வமான என் அவசரத்தையும் சேருங்கள். முதலில் வருத்தம், பின்னர் ஆதரவின்மை, அடுத்து முட்டாள்தனம் என்று அலைமேல் அலையாக உணர்ச்சிகளுக்கு ஆளானேன்.

தமிழ் மொழியைக் கற்றுக் கொள்வதில் ரொம்பவும் தினரும்போது கொஞ்சம் விட்டுவிடுகிறேன். ஆனால் நான் மறுபடியும் திரும்பி வருகிறேன். இது ஒரு வாழ்நாள் பயணமாக இருக்கலாம் என்ற உண்மையுடன் நான் என் மனதுக்குள் சமாதானம் ஆகிவிட்டேன். நான் வகுப்புகளுக்குச் செல்கிறேன், அதற்கான உழைப்பைப் போடுகிறேன், அழுகிறேன், திரும்பவும் தொடங்குகிறேன்.”

தாய்மொழியை இழந்தால் நாம் காட்டுவாழ்க்கையில் இருந்து மீண்டும் ஆரம்பிக்கிறோம்

எழுத்தாளர் ஜெயமோகன் கூறுகிறார், “பண்பாட்டால்தான் நம் ஆழ்மனம் உருவாக்கப்பட்டிருக்கிறது. நம் மனதின் வயது நம் வயதே. நம் ஆழ்மனத்தின் வயதுக்கு நாம் நம் தொல்பழங்குடிகள் வரை சென்ற மூலம் கண்டுபிடிக்க வேண்டும். பண்பாட்டையும் மொழியையும் இழப்பதென்பது அந்த ஆழத்தை ஒட்டுமொத்தமாக இழப்பது... மொழி என்பதை ஒரு தொடர்புறுத்தும் சாதனம் மட்டுமே என்று சொல்லும் இயந்திரவாதக் கண்ணோட்டத்தை விரிவாக மறுத்து இப்போது மிக விரிவாகவே மொழியியல் பேசிக்கொண்டிருக்கிறது.

மொழி என்பது பண்பாட்டின் புற வடிவம். மொழியை இழக்கும்போது நீங்கள் இழப்பது ஒரு வகை ஊடகத்தை அல்ல. ஒரு மாபெரும் தொடர்ச்சியை... அதை இழந்தால் நாம் காட்டுவாழ்க்கையில் இருந்து மீண்டும் ஆரம்பிக்கிறோம்... எப்படி வாழ்ந்தாலும் பல்லாயிரம் வருடத்து மரபின் நீட்சியாக உள்ள ஒரு பெரும் செல்வத்தை- நம் ஆழ்மனத்தை- அதில் உள்ள ஆழ்படிமங்களை- நாம் ஏன் இழக்க வேண்டும்?... நம்புங்கள், நாம் எவருக்கும் குறைந்தவர்கள் அல்ல. மாபெரும் நாகரீகங்களை உருவாக்கிய மாபெரும் பரம்பரையினர் நாம். அந்தத் தொடர்ச்சி நமக்குச் செல்வமே ஒழிய சுமை அல்ல.”

தாய்மொழியை இழந்தால் நாம் கலாச்சாரமற்ற இயந்திர மனிதர்கள் ஆகிறோம்

தாய்மொழியை இழந்தால் நாம் காட்டுவாழ்க்கையில் இருந்து மீண்டும் ஆரம்பிக்கிறோம் என்று ஜெயமோகன் கூறுகிறார். ஆனால் மொழியை இழந்த இவர்களிடம் இல்லாத நவீன சாதனங்களே கிடையாது. எனக்குத் தோன்றுகிறது இவர்களை இயந்திர மனிதர்கள் என்று சொல்வதுதான் அதைவிட நல்ல உபமானம் என்று.



இயந்திரப் பிள்ளை

மொழியும் பண்பாடும் பிரிக்க முடியாத வகையில் இணைக்கப்பட்டுள்ளன. இயல், இசை, நாடகம் போன்ற நம் கலாச்சாரத்தின் இழைகள் நம்மை ஒரு சமூகமாகப் பிணைக்கின்றன. இந்தக் கலாச்சாரம்தான் தனிமனிதர்களின் கூட்டத்தை ஒரு சமூகமாக்குகிறது. கலாச்சாரம் மொழியில் உள்ளார்ந்ததாக இருக்கிறது. மொழி பொதுவான அனுபவங்களையும் பின்னணியையும் தொடர்புபடுத்துகிறது. மேலும் இம்மாதிரி வரலாறுகள் நம் கலாச்சாரம் சார்ந்த சொற்களை உருவாக்குகின்றன. ஒரு குழுவில் பகிர்ந்த அனுபவங்களைப் பற்றிப் பேசும்போது எல்லோரும் சிரிக்கிறார்கள் ஆனால் உங்களுக்குப் புரியவில்லை. இது எப்போதாவது உங்களுக்கு நடந்திருக்கிறதா? இந்த உள் நகைச்சுவை உங்களை விலக்கி வைப்பதுபோல் உணர்ந்தீர்களா? கலாச்சாரத்தைப் புரிந்துகொள்ளாமல் சொற்களை மட்டும் புரிந்துகொள்வது கடினம். பகிர்ந்த அனுபவம் சொற்களின் பொருளைப் பாதிக்கிறது என்பதை இது காட்டுகிறது.

பிள்ளைகள் தமிழ் பேச புலம்பெயர்ந்த பெற்றோர்கள் என்ன செய்ய வேண்டும்?

பகல்நேரக் குழந்தைகள் கவனிப்பு மற்றும் மழலையர் பள்ளிக்குச் செல்ல ஆரம்பித்தவுடன் நாள் முழுவதும் மற்றக் குழந்தைகளுடன் ஆங்கிலத்தில் பேசுவதால் அதை எளிதில் கற்றுக் கொள்வார்கள். அதற்கு முன்னரே நீங்கள் வசிக்கும் பகுதியில் மற்றக் குழந்தைகளுடன் விளையாடும்போதே கற்றுக் கொள்ளவும் வாய்ப்புண்டு. வீட்டிற்குத் திரும்பிய பின்பும் அதே பழக்கம் தொடர்வது இயற்கைதான். ஆனால் தமிழை சொல்லித் தரும் பொறுப்பு உங்களுடையது. தமிழைப் பயன்படுத்துவதற்கு குழந்தைகள் தயக்கம் காட்டுவது பொதுவாக இந்த இரு முக்கியப் பகுதிகளில் உள்ள

குறைபாடுகளால்தான் – தேவை மற்றும் சுற்றுச்சூழல். ஆகவே பிறந்தது முதல் தமிழைக் கற்க வேண்டிய தேவையை ஏற்படுத்துவதுடன் எப்போதும் அவர்கள் காதுகளில் தமிழ் ஒலி கேட்டுக்கொண்டே இருக்கும் சுற்றுச்சூழலை ஏற்பாடு செய்வது உங்கள் பொறுப்பு. ஒரு குழந்தை பேசத் தொடங்குகையில் தமிழ் பேசத் தயங்கினால் அதற்கு இந்த இரண்டு காரணங்கள்தான்.

கருப்பையில் உள்ள குழந்தை 30 வாரத்தில் தொடங்கி தங்கள் தாய்மார்களிடமிருந்து மொழியைக் கற்கத் தொடங்குவதாக சமீபத்திய ஆராய்ச்சியில் தெரியவந்துள்ளது. பிறந்த குழந்தைகள் சில மணி நேரங்களிலேயே தங்கள் தாயின் மொழிக்கும் மற்ற மொழிக்கும் இடையில் வேறுபாடு அறிய இயலும். ஏனென்றால் அவர்கள் கருப்பையிலேயே தங்கள் அன்னையின் பேச்சு வடிவங்களையும் தனித்தன்மை வாய்ந்த ஒலிகளையும் கேட்டுக்கொண்டிருக்கிறார்கள்.

உங்கள் பிள்ளைகள் வளர்ந்தபின் அவர்கள் மேற்கண்ட மாதிரி சித்திரவதைக்கு உள்ளாகாமல் இருப்பது உங்கள் கையிலுள்ளது. மேலும் சாதிக்க விழையுங்கள் ஆனால் நீங்கள் யார் என்பதை மறந்துவிடவேண்டாம். ஆங்கிலத்தின் மூலம் உலகளாவிய தொடர்புகளுக்கும் நல்ல பணி வாய்ப்புகளுக்கும் விழையுங்கள், ஆனால் நம்முடைய கலாச்சாரத்துக்கு ஆணிவேரான தமிழை உங்கள் பிள்ளைகளுக்குக் கற்பிக்க மறந்து விடாதீர்கள்.

10. கணினிக்குத் தமிழ் கற்றுக் கொடுப்போம் வாருங்கள்

பண்டைய காலத் தமிழர் இலக்கியப் படைப்புகளை மனப்பாடம் செய்து காத்தனர்

இறையனார் களவியல் அல்லது அகப்பொருள் உரையை உருவாக்கியவர் நக்கீரர். இவரது காலம் கி.பி. 7-ம் நூற்றாண்டு வாக்கில். இவர் தாம் செய்த களவியல் உரையை வாய்மொழியாகத் தம் மகனார் கீரங்கொற்றனாருக்கு உரைத்தார். கீரம் கொற்றனார் தேனூர் கிழாருக்கு உரைத்தார். இவ்வாறாக இந்த உரை அடுத்தடுத்து மனப்பாடமாக எட்டு தலைமுறைக்கு ஒப்படைக்கப்பட்டது. இறுதியில் கி.பி. 10-ம் நூற்றாண்டு வாக்கில் முசிரி நீலகண்டன் இந்த உரையை ஓலைச்சுவடியில் எழுதி வைத்தார்.

அடுத்து அவை ஓலைச்சுவடியில் கையால் எழுதிப் பாதுகாக்கப்பட்டன

நீண்ட காலம் அழியாதிருக்க வேண்டும் எனக் கருதப்பட்ட செய்திகள் கற்களில் வெட்டப்பட்டன. பெரும்பாலும், மன்னர்களின் ஆணைகள், அவர்கள் செய்த பணிகள் போன்றவை கல்வெட்டுக்களாகப் பொறிக்கப்பட்டன. தென்னிந்திய அரச மரபினரால் வழங்கப்பட்ட பல கொடைகள் குறித்த பதிவுகள் செப்பேடுகளில் எழுதப்பட்டன. இலக்கியம், இலக்கணம் போன்ற நீண்ட உரைகள் பனையோலையில் எழுதப்பட்டன. இவற்றை எழுத்தோலை அல்லது ஓலைச்சுவடி என்று கூறுகிறோம்.

பின்னர் அச்ச இயந்திரம் மூலம் பல படிகள் எடுக்க வழி கிடைத்தது

உ. வே. சாமிநாதையர் பல தமிழ் இலக்கியங்களின் ஓலைச்சுவடிகளை மிகவும் மெனக்கெட்டு சேகரித்து அச்சில் வெளியிட்டார் என்பது பரவலாகத் தெரிந்ததுதான். ஆனால் சேலம் இராமசாமி முதலியார் "இதெல்லாம் படித்து என்ன பிரயோசனம்? இவைகளெல்லாம் பிற்கால நூல்கள், இவைகளுக்கு மூலமான நூல்களைக் கற்றுள்ளீர்களா?" என்று அவரைக் கேட்டு, சில மூல நூல்களையும் கொடுத்து, அவற்றின் பொருள் பற்றி அவரிடம் விவாதித்து அவர் பல சங்கத் தமிழ் இலக்கியங்களைக் கண்டுபிடித்து பதிப்பிக்க தூண்டுகோலாக இருந்தார் என்பது உ.வே.சா. சுயசரிதையைப் படித்தவர்களுக்கு மட்டும்தான் தெரியும். அவருக்குப் பல ஆண்டுகள் முன்னரே அரிய தொல்காப்பிய உரைகள், இறையனார் அகப்பொருள், கலித்தொகை போன்ற பல பண்டைய சங்கத் தமிழ் நூல்களை அச்சிட்டு வாழ வைத்த தமிழ்ப் பதிப்புத்துறையின் முன்னோடி யாழ்ப்பாணம் சி. வை. தாமோதரம்பிள்ளை பற்றி பலருக்குத் தெரியாது.

அடுத்து இணையத்தில் மின்தொகுப்பாக்கிப் பகிர்கிறோம்

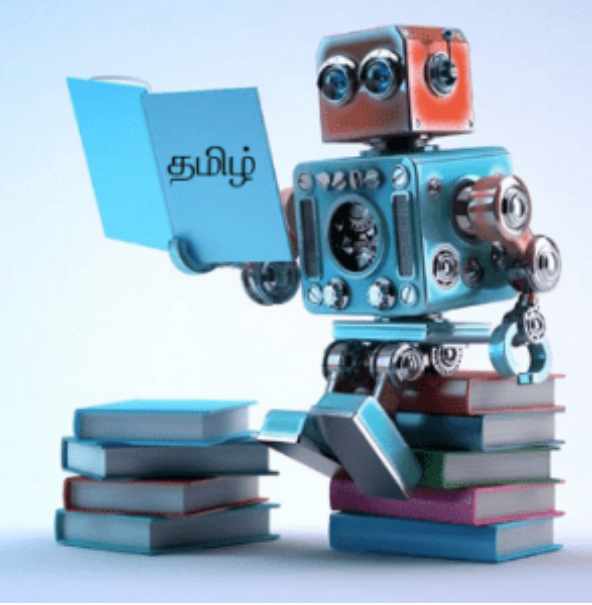
பண்டைய இலக்கியப் படைப்புகளை இணையத்தில் அனைவரும் அணுக இயலும் மின் வடிவத்தில் மாற்றும் முயற்சிகள் பல உலகளவில் முன்னெடுக்கப்பட்டன. முனைவர் கு. கல்யாணசுந்தரம் தலைமையில் நடக்கும் மதுரை தமிழ் இலக்கிய மின்தொகுப்புத் திட்டம் இவற்றில் முக்கியமானதொன்றாகும். நூலகம் திட்டம் சுமார் 7000 நூல்களையும் மற்றும் பல்லாயிரக் கணக்கான இதழ்கள், செய்தித்தாள்கள் ஆகியவற்றையும் மின் தொகுப்புகளாகப் பதிவேற்றி உள்ளது. தமிழ் இணையக் கல்விக்கழகம் பல சங்க இலக்கியங்கள், காப்பியங்கள், நெறி நூல்கள் முதலியவற்றை இணையதளத்தில் அளிக்கிறது. தற்போது விக்கிப்பீடியாவின் விக்கிமூலம் தளத்தில் தமிழக அரசின் நாட்டுடைமையாக்கப்பட்ட நூல்கள் பதிவேற்றத் திட்டம் நடந்து கொண்டிருக்கிறது. 91 ஆசிரியர்களின் 2217 நூல்கள் எழுத்துணரி மூலம் பதிவேற்றப்பட்டு தன்னார்வர்களால் கூட்டு முயற்சியாக மெய்ப்புப் பார்க்கப்படுகிறது. ஆர்வமுள்ளவர்கள் முகப்புப் பக்கத்திலிருக்கும் இந்த மாதத்தின் புத்தகத்திற்குப் பங்களிக்கலாம். தமிழில் மின்னூல்களை இலவசமாக வெளியிடும் இந்தத் தளம் பல எழுத்தாளர்கள் தங்கள் படைப்புகளை வெளியீடு செய்ய உதவியிருக்கிறது. 400 நூல்கள் மொத்தம் சுமார் 6 மில்லியன் பதிவிறக்கம் செய்யப்பட்டுள்ளன.

இயல்மொழி ஆய்வுதான் தமிழின் பரிணாம வளர்ச்சியின் அடுத்த கட்டம்

பல்லாயிரக்கணக்கான ஆவணங்கள் இணையத்திலும், சமூக ஊடகங்களிலும் தினமும் சேர்கின்றன. இவற்றையெல்லாம் ஒவ்வொன்றாகப் படித்துப் பிரித்தெடுப்பதென்பது ஆகக்கூடிய வேலையில்லை. இதேபோன்று வானொலியிலும், தொலைக்காட்சியிலும், யூடியூபிலும் பல்லாயிரக்கணக்கான சொற்பொழிவுகளும், உரையாடல்களும், நேர்காணல்களும் வந்துகொண்டேயிருக்கின்றன. நிறுவனங்களும், கல்வியாளர்களும், அரசாங்கமும், நீங்களும், நானும் இவற்றைப் பயன்படுத்த வேண்டுமானால் இவற்றை வகைப்படுத்தவும், திறவுச்சொல் பிரித்தெடுக்கவும், சுருக்கம் தயாரிக்கவும் கணினியால் இவற்றைப் பகுப்பாய்வு செய்ய இயலவேண்டும். இதற்கு இயல்மொழிக் கருவிகளும் வளங்களும் தேவை.

நம்முடைய இயல்மொழியைப் புரிந்து கொள்ளக் கணினிகளைப் பழக்கி விட முடியும்

இது ஒரு புறமிருக்க, இதுநாள்வரை கணினிகளைப் பயன்படுத்துவது எப்படி என்று நாம் கற்றுக் கொள்ள வேண்டியிருந்தது. ஆனால் செயற்கை நுண்ணறிவு மற்றும் இயந்திரக் கற்றலின் சாத்தியம் என்னவென்றால் நம்முடைய இயல்மொழியைப் புரிந்து கொள்ளக் கணினிகளைப் பழக்கி விட முடியும் என்பதுதான். இதுநாள்வரை கணினிகளைப் பயின்றவர்கள்தான் அவற்றின் திறனைப் பயன்படுத்த முடியும் என்றிருக்கிறது. நம்முடைய இயல்மொழியைப் புரிந்து கொள்ளக் கணினிகளைப் பழக்கி விட்டால் பாமர மக்களும் கணினிகளின் திறனை நேரடியாக அணுக இயலும். இதற்கும் இயல்மொழிக் கருவிகளும் வளங்களும்தான் அடிப்படைத் தேவைகளாக அமைகின்றன.



கணினிக்குத் தமிழ் கற்றுக் கொடுப்போம்

இயற்கையான மொழியைப் பயன்படுத்திக் கணினிகள் மற்றும் பிற சாதனங்களை மனிதர்கள் எளிதாக இயக்குவதற்கான செயலிகளை உருவாக்குவதுதான் இயல்மொழி ஆய்வின் குறிக்கோள். எடுத்துக்காட்டாக ஆங்கிலத்தில் கணினிகளுடன் உரை அரட்டையும் பேச்சு உரையாடலும் வந்து விட்டது. கூகிள் போன்ற பெரிய நிறுவனங்கள் தமிழிலும் சில சேவைகளைத் தொடங்கியுள்ளன. இந்தக் குறிக்கோளை அடையும் முதல் படியாக உரை மற்றும் பேச்சு போன்ற இயற்கையான மொழிகளில் உட்பொதிக்கப்பட்ட தகவல்களை கணினி மூலம் பிரித்தெடுக்க வேண்டும்.

மற்றும் கணினியில் மொழிபெயர்ப்பு முன்னேற்றங்கள் வரவர ஒரு பொதுவான மொழியின் தேவை குறைந்துவிடும். உங்கள் உலாவி உங்கள் மொழியில் தானியங்கியாக மொழிபெயர்ப்பு செய்யமுடியும் என்றால், அல்லது உங்கள் திறன்பேசி மொழிபெயர்ப்பாளராக இயங்க முடியும் என்றால் ஏன் இன்னொரு மொழியைக் கற்க முயற்சி போடுவோம்?

செயற்கை நுண்ணறிவும் இயந்திரக் கற்றலும் ஆழ்ந்த கற்றலும்

இயந்திரக் கற்றல் என்பது செயற்கை நுண்ணறிவின் ஒரு பகுதிக்கு வழங்கப்படும் பெயர். இது தரவுகளைப் பார்த்துக் கற்றுக் கொள்ளும் மற்றும் மேம்படுத்தும் வினைச்சரங்கள் சம்பந்தப்பட்டது. இந்த இயந்திரக் கற்றல் வினைச்சரத்துக்கு நாம் இயல்மொழி தரவுகளை பெரும்பாலும் உரை வடிவத்தில் கொடுக்கிறோம். இந்த உரை நாம் செய்யும் கற்றல் பணிக்குத் தொடர்புடைய குறிப்பிட்ட அம்சங்களை முன்னிலைப்படுத்துமாறு குறியீடு செய்யப்பட்டுமிருக்கலாம்.

இயல்மொழி ஆய்வுத் துறையில் தரவுத் தொகுப்புகளுடன் பணி புரியும்போது, பொதுவாக மூன்று முக்கிய வகையான இயந்திரக் கற்றல் நெறிமுறைகள் பயன்படுத்தப்படுகின்றன. கையால் குறியீடு செய்த உரையைப் பயன்படுத்திக் கற்பித்தலை நாம் மேற்பார்வையில் கற்றல் என்று சொல்கிறோம். குறியீடு செய்யாத உரையைப் பயன்படுத்திக் கற்பித்தலை மேற்பார்வையற்ற கற்றல் என்றும் இரண்டும் கலந்தது என்றால் அரை மேற்பார்வை என்றும் சொல்கிறோம்.

தொகுப்பில் இருக்கும் ஒரு உருபொருளை அடையாளம் காண்பதை வகைப்படுத்தல் என்கிறோம். எடுத்துக்காட்டாக, உங்கள் மின்னஞ்சல் அகப்பெட்டியில் வீண்செய்திகளை வேறுபடுத்துவதற்கு வகைப்படுத்தல் வினைச்சரம் பயன்படுகிறது. ஜிமெயில், யாஹூ, ஹாட்மெயில் போன்றவற்றில் உங்களுக்கு வரும் மின்னஞ்சல்களில் எது வீண்செய்தி என்று அடையாளம் செய்தால் அதிலுள்ள சக்தி வாய்ந்த வகைப்படுத்தல் வினைச்சரம் உடன் கற்றுக் கொள்ளும்.

நீங்கள் ஐந்து ஆவணங்களைப் படித்து விட்டு உரைச் சுருக்கம் எழுதக்கூடும். ஐயாயிரம் ஆவணங்கள் என்றால்? ஒவ்வொரு ஆவணத்திலும் இருக்கும் மையக்கருத்துக்கள் மற்றும் முக்கிய சொற்றொடர்கள் தெரிந்தால் அவற்றை பயனர்கள் தானியங்கியாக அடையாளமிடவும், வகைப்படுத்தவும், முறைப்படுத்தவும் இயலும். இது ஆய்வாளர்களுக்கும் தரவுத்தள மேலாளர்களுக்கும் மிகவும் உதவியாக இருக்கும். ஆவணங்களில் மையக்கருத்துக்கள் மற்றும் முக்கிய சொற்றொடர்களைக் கணினி மூலம் கண்டுபிடிப்பதைத் தலைப்புப் பிரித்தெடுத்தல் (topic extraction) என்கிறோம்.

ஆழ்ந்த கற்றல் மேலும் மனித மூளையை ஒத்திருக்கிறது. மூளை மண்டலத்தின் ஆழத்தில் உள்ள நரம்பணுக்கள் போல நுணுக்கமான கட்டமைப்புகளை அடையாளம் காண்கிறது. இவற்றைப் பற்றி விரிவாகப் பின்வரும் கட்டுரைகளில் காண்போம்.

இது பார்வையாளர் விளையாட்டல்ல, சட்டைக் கைகளை மடக்கி விட்டுக் கொண்டு களத்தில் இறங்குங்கள்

அனைத்து மொழி பேசுவோரும் சமமான நிலைப்பாட்டில் பங்கேற்பதை உறுதி செய்ய உரை மற்றும் பேச்சுத் தொழில்நுட்பம் பயன்படுத்தப்படும். ஆங்கிலம், ஸ்பானிஷ், பிரஞ்சு, ஜெர்மன் போன்ற வர்த்தக ரீதியாக முக்கியமான மொழிகளைப் பற்றிக் கவலைப்பட வேண்டியதில்லை என்பது தெளிவாகத் தெரிகிறது. இந்த மொழிகளைப் பொறுத்தவரை ஒரு பெரிய செல்வந்த சந்தை சாத்தியமானது. அதனால் பெரிய சந்தை நிறுவனங்கள் இந்த மொழிகளைக் கவனித்துக்கொள்வார்கள். சிறிய மொழிகளில் சிறிய தேசிய, வட்டார பொருளாதாரங்கள்தானே உள்ளன. ஆனால் சந்தை அதைச் செய்யாவிட்டால் தமிழ் மொழிக்கு முறையான தொழில்நுட்பங்களை யார் வழங்குவார்? தீர்வு: சமூக முயற்சி, தன்னார்வலர்கள், திறந்த மூலம், திறந்த தரவுகள், திறந்த ஆய்வு. இது பார்வையாளர் வேடிக்கை பார்க்கும் விளையாட்டல்ல, சட்டைக் கைகளை மடக்கி விட்டுக் கொண்டு களத்தில் இறங்குங்கள்.

11. பெரு நிறுவனங்களின் தமிழ் சேவைகளை நம்பியே இருந்தால் என்ன?

கூகிள், ஆப்பிள், முகநூல், அமேசான், மைக்ரோசாப்ட் போன்ற பெரு நிறுவனங்கள் மொழி பற்றிய சேவைகள் பலவற்றை இலவசமாகத் தருகின்றன. எடுத்துக்காட்டாக கூகிள் தரும் மொழிக் கருவிகளைப் பற்றி விவரமாகப் பார்ப்போம்.

கூகிள் ஜி-போர்ட் – தமிழில் தட்டச்சும் சொல்வதெழுதலும்

ஜி-போர்ட் என்பது ஆண்ட்ராய்ட் மற்றும் ஆப்பிள் சாதனங்களுக்காக கூகிள் உருவாக்கிய மெய்நிகர் விசைப்பலகை செயலியாகும். தற்போது ஆண்ட்ராய்ட் திறன்பேசிகளில் 300 க்கும் மேற்பட்ட மொழிகளை ஆதரிக்கிறது. இதன் கணிக்கக்கூடிய இயந்திரம் (predictive typing engine), நாம் தட்டச்சு செய்து கொண்டிருக்கும் போது, சற்று முன் உள்ளிட்ட சொற்றொடர்களை வைத்து அடுத்த சொல் என்னவாக இருக்கும் என்று ஊகித்து சில சொற்களை யோசனை சொல்கிறது. அவற்றில் ஒன்று சரியாக இருந்தால் நாம் ஒவ்வொரு எழுத்தாக அடிக்காமல் உடன் தேர்வு செய்யலாம்.

இதன் பிப்ரவரி 2017 வெளியீடு சொல்வதெழுதல் ஆதரவைச் சேர்த்துள்ளது. பயனர்கள் மைக் பொத்தானை அழுத்திப் பேசினால் தானே எழுதி விடும். தவறுகளை மட்டும் திருத்திக் கொள்ளலாம். ஏப்ரல் 2017 வெளியீடு தமிழைச் சேர்த்து, மொத்தம் 22 இந்திய மொழிகளை ஆதரிக்கிறது. ஒற்றை வார்த்தைகளை விட முழுச் சொற்றொடரை ஊகிப்பதற்காக ஜூன் 2017 இல் மேம்படுத்தப்பட்டுள்ளது. ஆனால் இந்த வசதி இன்னும் தமிழில் வரக் காணோம்.

கூகிள் ஒளி எழுத்துணரி

கூகிள் ஒளி எழுத்துணரி (Optical Character Recognition OCR) மென்பொருள் இப்போது 248 க்கும் மேற்பட்ட உலக மொழிகளுக்கு (அனைத்து முக்கிய தென் ஆசிய மொழிகள் உட்பட) வேலை செய்கிறது. இது படங்களிலிருந்து உரைகளைப் பிரித்தெடுக்கிறது. இது அச்சிடப்பட்ட உரைகள், கையெழுத்துப் போன்றவற்றைக் கூட உரையாக மாற்றும். எனவே, பழைய புத்தகங்களிலோ, கையெழுத்துப் பிரதிகளிலோ அல்லது படங்களிலோ இருக்கும் உரையைப் பிரித்தெடுக்கலாம்.

[கூகிள் ஒளி எழுத்துணரியைப் பயன்படுத்த](#) எளிதான வழி கூகிள் ஆவணங்கள் (Google Docs) மூலமாகத்தான். பிரித்தெடுக்க வேண்டியதை JPEG, PNG, GIF, அல்லது பல பக்கங்கள் கொண்ட PDF கோப்பாக கூகிள் ஆவணங்களுக்கு பதிவேற்றுங்கள். பட்டியலில் அந்த கோப்பை வலச் சொடுக்கு செய்து Open With -> Google Docs தேர்வு செய்யுங்கள். கூகிள் ஒளி எழுத்துணரி தானே அது எந்த மொழியில் எழுதப்பட்டுள்ளது என்று கண்டுபிடித்து உரையாக மாற்றும்.

கூகிள் மொழிபெயர்ப்புச் செயலி

ஆண்ட்ராய்ட் மற்றும் ஆப்பிளுக்கான கூகிள் மொழிபெயர்ப்பு செயலி 100 க்கும் மேற்பட்ட மொழிகளுக்கு ஆதரவளிக்கிறது. இரண்டு மொழிகளுக்கு இடையே இயந்திர மொழிபெயர்ப்பு அமைப்பை உருவாக்க 150-200 மில்லியன் சொற்கள் கொண்ட இருமொழி சொல்வங்கியும் (bilingual text corpus), அவ்விரு மொழிகளுக்கும் தனித்தனியே இரண்டு பில்லியன் சொல்வங்கிகளும் (monolingual corpora) தேவைப்படும் என்று கூகிள் தொழில்நுட்ப வல்லுநர் கூறுகிறார். இந்தத் தரவுகளிலிருந்து புள்ளிவிவர மாதிரிகள் அந்த மொழிகளுக்கு இடையே மொழிபெயர்க்கப் பயன்படுத்தப்படுகின்றன.

கூகிள் மொழிபெயர்ப்பு பெரும்பாலும் ஒரு மொழியிலிருந்து மற்றொரு மொழிக்கு நேரடியாக மொழிபெயர்ப்பதில்லை. எடுத்துக்காட்டாக, நீங்கள் தமிழிலிருந்து வங்காளத்துக்கு மொழிபெயர்த்தால் அது உள்ளுக்குள் முதலில் ஆங்கிலத்திற்கு மொழிபெயர்த்து பின்னர் ஆங்கிலத்தில் இருந்து வங்காள மொழியில் மொழிபெயர்த்து முடிவை உங்களுக்குக் காட்டும்.

2017 மார்ச் மாதம் நரம்பியல் அடிப்படையிலான இயந்திர மொழிபெயர்ப்பை (Neural Machine Translation – NMT) கூகிள் மொழிபெயர்ப்பு அறிமுகப்படுத்தியது. முன்னர் இது சொற்றொடர் அடிப்படையிலான (Phrase Based) இயந்திர மொழிபெயர்ப்பாக இருந்தது. இந்தப் புதிய NMT முழு வாக்கியத்தையும் ஒரு உள்ளீடாகக் கருதி அதை முழுமையாக மொழிபெயர்க்கிறது. இது ஆழமான நரம்பியல் அடிப்படையிலான பிணையங்களைப் பயன்படுத்துகிறது. இது மற்ற தகவல்களிடமிருந்து கற்பதன் மூலம் ஒரு கணினியால் போகப் போக முன்பு பார்த்திராத சூழ்நிலைகளைப் புரிந்து கொள்ள உதவுகிறது. இக்காரணத்தால் 2017 மே மாதத்தில் கூகிள் வெளியிட்ட ஆங்கிலம்-தமிழ் மொழிபெயர்ப்பில் வாக்கிய அமைப்பு முன்னை விட நன்றாக இருப்பதைக் காணலாம்.

இணையத் தொடர்பு இல்லாத போதும் மொழிபெயர்க்கலாம். ஆனால் இதற்கு முன்னால் இணையத் தொடர்பு இருக்கும் இடத்தில் மொழித்தொகுப்பை நீங்கள் பதிவிறக்கி வைத்திருக்க வேண்டும். ஆனால் மற்ற செயலிகளிலிருந்து நிரலாக்க இடைமுகம் (API) மூலம் அனுப்பி மொழிபெயர்ப்பை வாங்க வேண்டுமென்றால் கட்டணம் உண்டு.



கூகிள் சொல் வில்லை

கூகிள் சொல் வில்லை

மொழிபெயர்ப்பி செயலியைப் பயன்படுத்தும் போது, பயனர்கள் தங்கள் நிழற்படக் கருவியை வேற்று மொழியிலுள்ள ஒரு சாலைக் குறியீட்டிலோ அல்லது உணவுப் பட்டியல் போன்ற உரையிலோ சுட்டிக்காட்டலாம். இணைய இணைப்பு இல்லாதபோதிலும் மொழிபெயர்க்கப்பட்ட உரையைப் பார்க்க முடியும்.

கூகிள் உரையாடல் செயலி

பல்வேறு மொழிகளில் உரையாடல்களைப் புரிந்துகொள்வதற்கான ஒரு மொழிபெயர்ப்புக் கருவியை கூகிள் அறிமுகப்படுத்தியுள்ளது. கூகிள் மொழியாக்கம் செயலியில், உரையாடல் முறையில் நுழைய வேண்டுமானால் பயனர்கள் மைக்கைத் தட்டலாம். இது உரையாடலில் இருவரும் என்ன மொழியில் பேசுகிறார்கள் என்று தானாகவே புரிந்து கொள்ளும். எடுத்துக்காட்டாக, நீங்கள் சுற்றுலா போன இடத்தில் சந்தையில் பேரம் பேச வேண்டுமானால், இந்த உரையாடல் முறை உங்களுக்கு உள்ளூர் மக்களுடன் உரையாட உதவும். இது முதலில் இந்திக்கு வேலை செய்தது. 2017 செப்டம்பரில் வங்காளம் மற்றும் தமிழ் மொழிகளைச் சேர்த்தார்கள்.

கூகிள் உரை ஒலி மாற்றி

கூகிள் உரை ஒலி மாற்றி என்பது ஆண்ட்ராய்டுக்காக அவர்கள் உருவாக்கிய திரை படிப்பி செயலி ஆகும். இது திரையில் இருக்கும் உரையை உரக்கப் படிக்கும். தற்போது இந்திய மொழிகளில் தமிழ், இந்தி மற்றும் வங்காளம் உட்பட கூகிள் உரை ஒலி மாற்றி 41 மொழிகளை ஆதரிக்கிறது. தமிழ் இன்னும் வரக்காணோம்.

கூகிள் உதவியாளர்

கூகிள் உதவியாளர் என்பது அவர்கள் உருவாக்கிய ஒரு தனிப்பட்ட மெய்நிகர் உதவியாளர். இது கை சாதனங்கள் மற்றும் திறன் வீட்டுச் சாதனங்களில் கிடைக்கிறது. இதனுடன் நீங்கள் இருவழி உரையாடல்களில் ஈடுபடலாம். இந்தி உட்பட பத்து மொழிகளில் கூகிள் உதவியாளர் கிடைக்கிறது, மேலும் ஆண்டின் இறுதியில் இது 30 க்கும் மேற்பட்ட மொழிகளில் கிடைக்கும் என்று சொல்கிறார்கள்.

பெரிய நிறுவனங்களின் இலவசத் தமிழ் சேவைகளையே நம்பி இருப்பதில் என்ன பிரச்சினை?

பெரிய நிறுவனங்கள் சேவைகளை எப்போது வேண்டுமானாலும் நிறுத்தலாம் அல்லது உங்களுக்குப் பழக்கமாகி விட்ட பின் அதற்கு விலை வைத்து விற்பனை செய்ய ஆரம்பிக்கலாம். மேலும் மூல நிரலும் தரவுகளும் நமக்குக் கிடைக்காது என்பதால் தேவைக்குத் தகுந்தாற்போல் நாம் விருப்பமைவு செய்துகொள்ளவும் இயலாது.

நாம் இலவச சேவைகளைப் பயன்படுத்தும்போது பெரிய நிறுவனங்கள் நம்மைப்பற்றிய பல அந்தரங்கத் தரவுகளைச் சேகரிக்கின்றன. இந்த அந்தரங்கத் தரவுகளை வைத்து நமக்கு என்ன பொருட்கள், சேவைகள் தேவைப்படலாம் என்று பார்த்து அதற்கான விளம்பரங்களை நமக்குக் காட்டுவதுதான் இந்நிறுவனங்களின் நோக்கம். தங்கள் சேவைகளை மேலும் துல்லியமாகச் செய்வதும் மற்றொரு நோக்கம். எடுத்துக்காட்டாக, நாம் உணவகம் தேடும்போது, நாம் எந்த இடத்தில் இருக்கிறோம் என்று தெரிந்தால், அருகில் இருக்கும் உணவகங்களைக் காட்ட இயலும். சமீபத்தில் கேம்பிரிட்ஜ் அனலிடிகா என்ற நிறுவனம் 87 மில்லியன் பயனர்களின் அந்தரங்கத் தரவுகளை முகநூலிலிருந்து அனுமதியின்றி எடுத்துக் கொண்டதைப் பற்றி நீங்கள் படித்திருக்கக்கூடும். பிரிட்டனிலும் மற்ற நாடுகளிலும் தேர்தலில் வேட்பாளர்களுக்கு ஆலோசனை கூற எடுத்ததாகத் தெரிகிறது. எல்லாப் பயனர்களின் பெயர்கள், பிறந்த தேதி, மின்னஞ்சல் முகவரி, கடவுச்சொல் மற்றும் பாதுகாப்பு கேள்விகளும் பதில்களும் திருடப்பட்டதாக யாஹூ முன்னர் அறிவித்து உங்களுக்கு ஞாபகம் இருக்கலாம். ஈக்விஃபாக்ஸ், யூபர் நிறுவனங்களும் இம்மாதிரி பெரிய அளவில் தரவுகளை இழந்துள்ளன. யாவரும் கடவுச்சொல்லை மாற்ற வேண்டும் என்று சமீபத்தில் ட்விட்டர் அறிவித்துள்ளது.

இருப்பினும் நாம் தொடர்ந்து நம் வாழ்க்கையை எளிதாக்கும் இலவச தளங்களில் ஈடுபடுகிறோம். எவ்வாறாயினும், தொழில்நுட்பம் முன்னேறும்போது, நாம் வரக்கூடிய சங்கடங்கள் பற்றி ஏதும் கவலையற்றிருந்தால், முன்னெப்போதையும்விட நமக்கு இடர்ப்பாடுகளுக்கு பாதுகாப்பே இல்லாமல் போகும்.

எல்லாவற்றையும் விட மேலாக, மொழிக்கான தொழில் நுட்பக் கட்டமைப்புகளும், தொழில் நுட்பக் கருவிகளும் பெருநிறுவனங்களின் தனிச்சொத்தாக இருப்பது மொழியின் வளர்ச்சிக்கும், மக்களுக்கும் நல்லதல்ல. ஏனெனில் இலாபத்தை அதிகமாக்குவது அவர்களுடைய நோக்கம். தமிழை வளர்ப்பதும் தமிழ் சமுதாயத்துக்கு எது நல்லதோ அதுவும்தான் நம் நோக்கம். இவை இரண்டும் எப்பொழுதுமே ஒருங்கிணைந்து வரும் என்று சொல்ல முடியாது.

12. ஏன் திறந்த மூலமும், திறந்த தரவுகளும், திறந்த ஆய்வும்?

இது நாள் வரை பொதுமக்களின் வரிப் பணத்தில் செய்யப்படும் ஆராய்ச்சிகளும், மென்பொருட்களும் பெரும்பாலும் சமூகம், பொதுமக்கள், அரசாங்கம் ஆகியவற்றுடன் தொடர்பில்லாமலே செய்யப்படுகின்றன. ஆராய்ச்சித் தரவும் மென்பொருட்களும் பெரும் செலவில் உருவாக்கப்படுகின்றன. ஆனால் அதைப் பகிர்ந்து கொள்வதேயில்லை. பெரும்பாலும் ஆராய்ச்சி முடிவுகள் வெளிவந்த பிறகு விரைவில் அந்தத் திட்டம் தரவுகளை இழந்து விடுகிறது. மென்பொருட்கள் மக்களுக்குப் பயன் தராமல் வீணாகின்றன. பண விரயம் மட்டுமல்லாமல், ஒத்துழைப்பையும் முன்னேற்றத்தையும் இந்த அணுகுமுறை தடுக்கிறது.

ஆய்வறிக்கைகள் படி இக்கருவிகள், வளங்கள் யாவும் உருவாக்கப்பட்டு விட்டனவே?

நீங்கள் கேட்கலாம், “ஆய்வறிக்கைகள் படி இக்கருவிகள் மற்றும் வளங்கள் யாவும் உருவாக்கப்பட்டு விட்டதாகத் தெரிகிறதே? முன்னர் கண்டுபிடித்த அதே சக்கரத்தை மீண்டும் கண்டுபிடிப்பதுபோல் அல்லவா இது இருக்கிறது? நாம் ஏன் மீண்டும் முயற்சி போடவேண்டும்?” என்று. ஆய்வறிக்கைகளைப் படித்தால் அப்படித்தான் தெரிகிறது. ஆனால் இவற்றை அணுகுவதோ வேறு சங்கதி. ஒன்றிரண்டு கருவிகள் பதிவிறக்கம் செய்யலாம். சில எழுதிக் கேட்டால் கிடைக்கக்கூடும். ஆனால் மொழிமாற்றிய (compiled) வடிவத்தில்தான் கிடைக்கும், மூல நிரல் (source code) கிடைக்காது. மூல நிரல் கிடைக்காவிட்டால் தேவைக்குத் தகுந்தாற்போல் மாற்றவோ மேம்படுத்தவோ முடியாது. தரவுகள் பெரும்பாலும் கிடைப்பதேயில்லை.

திறந்த ஆய்வும் திறந்த தரவுகளும்

ஆராய்ச்சி முடிவில் பொதுவாக ஒரு அறிக்கைதான் வெளியிடப்படுகிறது. அறிக்கையில் ஒருசில மாதிரி புள்ளிவிவரங்களும் இருக்கலாம். ஆனால் அந்த ஆராய்ச்சியின் முடிவுக்குக் கொண்டு செல்லும் மூலப் பொருள்களான தரவுகள் வெளியிடப்படுவதில்லை. இம்மாதிரி ஆராய்ச்சித் தரவுகள் பொதுவாகப் புள்ளிவிவரங்கள் மற்றும் எண்வகை போன்ற அளவீட்டுத் தரவுகளாக இருக்கலாம். அல்லது பேட்டி எழுத்துப்படிகள், எண்ணிம் உள்ளடக்கமான படங்கள், ஒலிப்பதிவு மற்றும் நிழல் படம் போன்ற பண்பு சார்ந்த தரவு வடிவத்திலும் இருக்கலாம். இத்தரவுகள் ஒரு முக்கியமான ஆதாரம் மட்டுமல்ல, கல்வி சார்ந்த முன்னேற்றத்திற்கும் மிக அவசியமானதாகும்.

அறிவியல் வெளியீடுகள் மற்றும் ஆராய்ச்சித் தரவுகளுக்கான திறந்த அணுகல் பற்றிய ஐரோப்பிய ஒன்றிய வழிகாட்டுதல்கள் இவ்வாறு கூறுகிறது. “பொது மக்களின் வரிப் பணத்தை ஏற்கெனவே ஒரு முறை செலவு செய்து சேகரித்த தகவலை ஒவ்வொரு முறையும் பொதுமக்களும் மற்ற ஆய்வாளர்களும் அணுகவும் அல்லது பயன்படுத்தவும் இயல வேண்டும். மேலும் அது ஐரோப்பிய நிறுவனங்களுக்கும், குடிமக்களுக்கும் முழுப் பயன் தர வேண்டும்.”

இந்தியாவிலும் மத்திய அரசின் அறிவியல் மற்றும் தொழில்நுட்பத் துறை, திறந்த அணுகல் என்ற இணைய தளத்தை உருவாக்கி ஆய்வுத் திட்டங்களின் தரவுகள் யாவற்றையும் திறந்த தரவுகளாக வெளியிட வேண்டுமென்று அறிவுறுத்தி வருகிறது.

திறமையாளர்களை ஈர்க்க திறந்த மூலம் வழி செய்கிறது

இன்னும் தீர்க்கப்படாத பிரச்சினைகளில் வேலை செய்யவே உருவாக்குநர்கள் விரும்புகிறார்கள். நீங்கள் எதிர்கொள்ளும் சுவாரசியமான சவால்களையும் மற்றும் அவற்றை தீர்வு செய்வது பற்றி நீங்கள் எப்படி சிந்தனை செய்கிறீர்கள் என்பதையும் நிரலாளர் சமூகத்துக்கு வெளிப்படுத்த திறந்த மூலம் உதவுகிறது. திறந்த மூல திட்டத்துக்கு திறமையுள்ள நிரலாளர்கள் எளிதில் பங்களிக்க இயலும். ஏற்கெனவே தீர்க்கப்பட்ட பிரச்சினைகளை அடிப்படையாக வைத்து, புதிய கண்டுபிடிப்புகள் உருவாக்க வழி வகுக்கிறது. தங்கள் சொந்த கள அறிவுக்கு அப்பால் மற்ற நிபுணர்களின் திறனையும் வைத்து நிரலாளர்கள் புதிய கண்டுபிடிப்புகள் உருவாக்குவதும் சாத்தியமாகிறது.

Select Language
Tamil

Select Gender
Male

IndicTTS

Sentence	HTS	HTS+STRAIGHT
அது தஞ்சாவூர்க் கோட்டைக்குள் பிரவேசிக்கவும் சக்கரவர்த்தியைப் பார்க்கவும் பயன்படலாம்.	▶	▶
அதற்குத் தகுந்தபடி ஏதாவது கொஞ்சம் பேசி வேஷம் போட்டால் போகிறது.	▶	▶
ஆனால் அவன் எதிர்பார்த்த சந்தர்ப்பம் ஒன்றும் கிட்டவில்லை.	▶	▶

இந்திய மொழிகளுக்கான உரை ஒலி மாற்றி

நடைமுறை எடுத்துக்காட்டு: இந்திய தொழில்நுட்பக் கழகத்தின் உரை ஒலி மாற்றி

[2010 இல் இந்திய தொழில்நுட்பக் கழகம் ஒரு அருமையான உரை ஒலி மாற்றியை உருவாக்கியது.](#) இது பெங்காலி, இந்தி, தமிழ், தெலுங்கு, மலையாளம் மற்றும் மராத்தி ஆகிய இந்திய மொழிகளுக்கான உரை ஒலி மாற்றி. தொகுப்புகளை உருவாக்குதல், தரத்தை மேம்படுத்துதல் மற்றும் செயலியின் கால் தடத்தைச் சுருக்கி ஊனமுற்றவர்களுக்கு உதவும் பிற செயலிகளுடன் ஒருங்கிணைத்தல் ஆகியவை இத்திட்டத்தின் குறிக்கோள். இது ஃபெஸ்டிவல் (Festival) என்ற திறந்த மூல மென்பொருளை அடிப்படையாகக் கொண்டது. 2012 ஆம் ஆண்டில் 13 இந்திய மொழிகளுக்கு HTS புள்ளியியல் உரை ஒலி மாற்றி அடிப்படையிலான திட்டத்தின் இரண்டாம் கட்டம் துவங்கியது. இந்தத் திட்டத்தில் சென்னை எஸ். எஸ். என். பொறியியல் கல்லூரியும் பங்களித்திருந்தது. இது நன்றாக வேலை செய்வதை நிரூபிக்க ஒரு இணைய தளம் இருக்கிறது. ஆனால் இதன் பயன் மக்களைச் சென்றடைந்ததா என்று தெரியவில்லை. கூகிள் உரை ஒலி மாற்றியில் இன்னும் தமிழ் வரக்காணோம். ஆக இந்தவொரு தொழில்நுட்பத்தில் கூகிளைவிட நம் திறந்த மூலம் ஒரு படி மேலே இருக்கிறது.

2017 இல் திரு. சீனிவாசன் அவர்கள் தகவல் உரிமைச் சட்டத்தின் கீழ் இந்த மென்பொருளை திறந்த மூலமாக வெளியிடக் கோரி ஒரு விண்ணப்பம் அனுப்பினார். [இந்திய தொழில்நுட்பக் கழகத்திலிருந்து உடன் பதில் கிடைத்தது, அது ஏற்கெனவே திறந்த மூலத்தில் பகிரப்பட்டிருப்பதாக! திரு. சீனிவாசன், திரு. மோகன் இருவரும் மிகவும் முயன்று இ.தொ.க. ஆய்வாளர்கள் உதவியுடன் திறந்த மூல ஃபெஸ்டிவல் உரை ஒலி மாற்றி கருவியில் இயங்கும் குரல் மாதிரிகளைத் தொகுக்க முடிந்தது.](#)

ஆனால் இதில் இன்னும் சில முன்னேற்றங்கள் தேவைப்பட்டன. இந்த செயலியால் அடுத்த வெற்று வரி வரைதான் படிக்க முடிந்தது. ஒரு முழுப் பக்கத்தையோ அல்லது முழுப் புத்தகத்தையோ படிக்குமாறு செய்ய வேண்டும். [அடுத்து திரு. மோகன் இந்தத் தமிழ் உரை ஒலி மாற்றியின் நிரலில் ஒரே நேரத்தில் பல உரைகளை மாற்றுவதை எளிதாக்கி உதவினார்.](#) இது உரையை ஒலியாக்கி .wav அல்லது .mp3 வடிவத்தில் வெளியிடும். அடுத்து இதை இணைய செயலியாக மாற்ற வேண்டும். அப்பொழுதுதான் இதை எவரும் எளிதில் பயன்படுத்த முடியும்.

திரு. லெனின் குருசாமி தலைமையில் காரைக்குடி க்னூ லினக்ஸ் பயனர் குழு இந்த உரை ஒலி மாற்றியைப் பயன்படுத்தி சில தமிழ்ப் புத்தகங்களை ஒலி வடிவத்தில் வெளியிட்டுள்ளனர். [இதற்கான இணையதளத்தை திரு. வேலுச்சாமி உருவாக்கியுள்ளார்.](#) இந்த ஒலிப் புத்தகங்கள் பார்வையற்றவர்கள் மற்றும் பார்வை குறைபாடுடையோருக்கு மிகப்பெரிய உதவியாக இருக்கும் என்பதில் ஐயமில்லை. [இந்த ஒலிப் புத்தகங்களை காப்பகம் இணைய தளத்திலும் பதிவேற்றியுள்ளனர்.](#)

இம்மாதிரி ஒரு தயாரிப்பையோ அல்லது செயலியையோ படிப்படியாக முன்னேற்றம் செய்வதும் மற்றும் தன்னார்வலர்கள் தம்மால் முடிந்ததை, முடிந்தபோது பங்களிப்பதும், பற்பல பயன்பாடுகளுக்குத் தகுந்தாற்போல் விருப்பமேவு செய்துகொள்வதும், இக்கூட்டு முயற்சிகளின் பலன் யாவருக்கும் எளிதில் அணுகக்கூடிய வகையில் கிடைப்பதும் திறந்த மூல திட்டங்களின் அடிப்படைப் பண்புகள்.

அடுத்த தலைமுறையை தயார் செய்தல்

மென்பொருள் தொழில்நுட்பத்தில் இன்று மிக முக்கிய பொறியாளர்களாக இருக்கும் பலரும் திறந்த மூலத்தில் வேலை செய்துதான் தங்கள் தொழில்நுட்ப அறிவைக் கூர்மைப்படுத்தினர். பார்வையிட வசதியாக மென்பொருளின் அடிப்படை நிரல் வெளியிடப்பட்டதால், தொழிற்சாலைகளின் அதிநவீன தொழில்நுட்பம் எப்படி கட்டப்பட்டுள்ளதென்று கணினி அறிவியல் கல்லூரிகளும் ஆய்வு செய்யலாம். அடுத்த தலைமுறை மென்பொருள் பொறியாளர்களை தயார் செய்வதில் இது பெரும் பங்காற்றும். திறந்த மூலம் இல்லாவிட்டால் மென்பொருள் உள்ளுக்குள் எப்படி வேலை செய்கிறதென்பதை வெறும் ஊகம்தான் செய்ய வேண்டியிருக்கும்.

மொழியியலாளர்களுக்கும் மென்பொருள் உருவாக்குநர்களுக்கும் இடையே நெருங்கிய ஒத்துழைப்பு தேவை

இத்துறையில் திறம்பட வேலைசெய்ய இரண்டு வேறுபட்ட துறைகளில் அறிவும் அனுபவமும் தேவை. ஒன்று மொழியியல், மற்றொன்று கணினி நிரலாக்கம். இவை இரண்டுமே ஒருங்கிணைந்த நிரலாக்க மொழியியல் பயிற்சிகள் உள்ளன. எனினும் திறந்த மூலமும், திறந்த தரவுகளும், திறந்த ஆய்வும்தான் துறைகள் அளவில் நெருங்கிய ஒத்துழைப்புக்கு வழிவகுக்கும்.

13. நிரல் எழுதத் தெரியாதவர்கள் 21 ஆம் நூற்றாண்டின் தற்குறிகளா?

ஒலைச்சுவடி காலத்திலும் காகிதம் வந்தவுடனும் தொழில் நெறிஞர்களே எழுத்தாளர்களாக பெரிய மனிதர்களின் ஆதரவில் பணியாற்றினர். இதன் விளைவாக எழுதப்படிக்கத் தெரிந்தவர்களுக்கு செல்வாக்கு அதிகரித்தது. பிள்ளைகளைப் படிக்க வைப்பது வாழ்க்கை முறையாயிற்று. பின்னர் எழுதப்படிக்க இயலாதவர்கள் தற்குறி எனப்பட்டனர்.

நிரலாக்கம்தான் புதிய எழுத்தறிவா?

ஏற்கனவே நிரலாக்கம் என்பது தொழில் ரீதியாக நிரல் எழுதுபவர்களுக்கு மட்டும்தான் என்றில்லாமல் எந்த உயர் கல்விக்கும் அத்தியாவசியம் என்றாகி விட்டது. நிரலாளர் அல்லாத ஆற்றல் மிக்க பயனர்கள் ஒரு வேலையைத் திரும்பத் திரும்பச் செய்ய வேண்டியிருந்தால் விரிதாள் அல்லது உரை செயலியில் துணுக்கு நிரல் எழுதி தானியங்கியாக ஓட வைப்பது பல ஆண்டுகளாகவே நடந்துகொண்டுதானிருக்கிறது. இது ஒன்றும் புதிதல்ல.

“நிரல் எழுத்தறிவு: நிரலாக்கம் எப்படி உரை எழுதுவதை மாற்றுகிறது” என்ற புத்தகத்தை அனெட் வீ எழுதியுள்ளார். தகவல்தொடர்புகள் மேலும்மேலும் நிரல்கள் மூலம் நடக்க ஆரம்பித்துவிட்டன என்று கூறுகிறார். கணினி அறிவு இல்லாதவர்கள் தங்கள் வாழ்க்கையை வழிநடத்த மற்றவர்களை நம்பியிருக்க வேண்டும். ஆங்கிலத்தில் பட்டம் பெற்றால்தான் ஒரு கடிதத்தையோ அல்லது அறிக்கையையோ ஆங்கிலத்தில் எழுதலாம் என்று இருக்க முடியுமா என்ன என்று கேட்கிறார்.

உயிரியலாளராக வேண்டுமா? நிரலாக்கம் நன்றாகக் கற்றுக்கொள்ள வேண்டும்!

“கடந்த வாரம் பாஸ்டன் சென்ற பொழுது உயிரியலில் முது முனைவர் படிக்கும் ஒரு சிலரைச் சந்தித்தேன். அவர்களில் எத்தனை பேர் நிரலாக்கம் கற்றுக் கொள்கிறீர்கள் என்று கேட்டேன். ஒட்டுமொத்தமாக அனைவரும் கையை உயர்த்தினார்கள். உயிரியலில் பெரிய தரவு அவசியமாகிவிட்டது. முன்னாலிருந்த சிறிய அளவிலான தரவுகளில் விரிதாள் பயன்படுத்திக் கண்டுபிடிக்க முடியும். ஆனால் கடந்த சில ஆண்டுகளில் தரவுகளின் அளவு அதிகமாகிவிட்டது. 15,000 தரவுப் புள்ளிகளை கைமுறையாகப் பார்க்க முடியாது. தங்கள் பரிசோதனைக்காக எல்லாவற்றையும் ஆய்வு செய்வதற்கு தனிப்பட்ட நிரல்களை எழுத வேண்டியிருக்கிறது.”

சான் பிரான்சிஸ்கோவில் உள்ள கலிஃபோர்னியா பல்கலைக்கழகத்தில் (UCSF), அனடோல் கிரீட்ஸர் புதிதாக துறைத் தலைவராகப் பதவியேற்றார். நரம்பியல் பட்டதாரி மாணவர்களுக்கு பாடத்திட்டத்தை சீரமைக்க முயல்கிறார். அவருடைய முதல் நடவடிக்கைகளில் ஒன்று நரம்பியல் முக்கிய பாடத்திட்டத்தில் நிரலாக்கத்தை இணைக்க சிறந்த வழியைக் கண்டுபிடிக்க ஒரு குழுவை அமைப்பதுதான்.

பென்சில்வேனியா பல்கலைக்கழகத்தில் கல்வி பேராசிரியரான யாசிம் கபாய், மற்றும் சார்லஸ்டன் கல்லூரியில் கல்வி பேராசிரியரான க்வின் பர்க் சொல்கிறார்கள், “இளைஞர்களுக்கு நிரலாக்கத்தை அறிமுகம் செய்யும் பாடங்கள் அவர்களை நிரலாளர்களாக ஆக்குவதற்கு அல்ல. ஆனால் அவர்களை எண்ணிம ஊடகத்தின் மிகவும் பயனுள்ள படைப்பாளிகள் மற்றும் நுகர்வோராக ஆக்குவதற்குத்தான்.”



நிரலாக்கச் சிந்தனை: முக்கிய கருத்துகள்

“நிரலாக்கச் சிந்தனை” என்றால் என்ன?

நிரல் வரிகளை ஆயிரக்கணக்கில் எழுதித் தள்ளுவதைவிட முக்கியமாக, நாம் முதலில் நிரலால் எம்மாதிரி பிரச்சினைகளுக்கு தீர்வு காண முடியும் என்று புரிந்து கொள்ள வேண்டும். அமெரிக்காவுக்கு நிரல் எழுதுங்கள் என்ற திட்டத்தில் இளம் நிரலாளர்கள் கொண்டுவரும் மிகப் பெரிய பங்களிப்பு அவர்கள் எழுதும் மென்பொருள் அல்ல. அவர்களுடைய வித்தியாசமான கண்ணோட்டம்தான். இதைத்தான் “நிரலாக்கச் சிந்தனை (computational thinking)” என்று சொல்கிறோம். நிரல் எழுதி நம்மைச் சுற்றியுள்ள பிரச்சினைகளுக்கு எப்படி தீர்வு காண்பது என்று உங்களால் யோசிக்க முடியாவிட்டால், உலகிலுள்ள அனைத்து ஜாவா தொடரியலையும் தெரிந்து கொண்டாலும் பயனில்லை.

நிரலாக்கச் சிந்தனையில் நான்கு முக்கிய கருத்துகள் உள்ளன. பிரச்சினையை சிறு துண்டுகளாக உடைப்பதை பிரித்து ஆய்வு (decomposition) என்கிறோம். பல்வேறு சங்கதிகளுக்கு இடையில் ஒற்றுமைகளைக் கண்டறிதலை உருவகை அறிதல் (pattern recognition or matching) என்கிறோம். குறிப்பிட்ட வேறுபாடுகளை ஒதுக்கிவிட்டு பல சிக்கல்களுக்கு ஒரே தீர்வை வேலை செய்ய வைப்பதை அருவமாகக் கருதல் (abstraction) என்கிறோம். ஒரு பணியை முடிக்க நாம் பின்பற்றும் படிகளின் பட்டியலை வினைச்சரம் (algorithm) என்கிறோம்.

முதலில் பிரச்சினை என்ன என்று தீர்க்கமாகப் புரிந்து கொள்வது, அடுத்து அதற்குத் தீர்வு எப்படிச் செய்வது என்று முடிவு செய்வது பின்னர் நிரல் எழுதுவது – இதுதான் “நிரலாக்கச் சிந்தனை”. இவற்றில் மூன்றாவதாக உள்ள நிரல் எழுதுவதற்குக்கூட மற்றவர்கள் உதவி செய்ய இயலும். முதல் இரண்டும்தான் நிரலாக்கச் சிந்தனைக்கு இன்றியமையாதவை.

இப்போது கணினி ஆதிக்கம் எல்லாத் தொழில் வாழ்க்கையிலும் ஊடுருவி விட்டதால் இது நிரலாளர்களுக்கு மட்டுமல்ல. நிரலாக்க அணுகுமுறை உலகளாவிய பிரச்சினைகளை புதிர்களாகப் பார்ப்பதில் வேருன்றியுள்ளது. இவற்றை சிறிய பாகங்களாகப் பிரித்து தர்க்கம் மற்றும் ஊகிக்கும் முறையால் கொஞ்சம் கொஞ்சமாகத் தீர்வு செய்யலாம். மகிழ்ச்சியான உண்மை என்னவென்றால், கணினிகள் எப்படி வேலை செய்கின்றன, எப்படி அவை புரிந்து கொள்ளும் மொழியில் ஆணையிடுவது என்று நாம் தெரிந்து கொண்டால், கணினி செய்யக்கூடிய ஒரு திட்டத்தை நாம் உருவகப் படுத்தலாம். அது ஒரு நிரலாளருக்குப் புரியும் விதத்தில் விவரிக்கலாம், விவாதிக்கலாம்.

ஏனென்றால் நிரலாளர்களே உங்களுக்குச் சொல்வார்கள், ஒரு திட்டத்தை உருவாக்குவது பெரும்பாலும் கடினமான பகுதி இல்லை, எதை உருவாக்குவது மற்றும் ஏன் என்பதுதான். கணினி அறிவியலர்கள் பல ஆண்டுகளாக எச்சரிக்கை விடுத்துக்கொண்டுள்ளனர் – நிரலைப் புரிந்துகொள்வதும், எழுதுவதும் ஒரு நாள் இன்றியமையாததாக ஆகிவிடும் என்று.

கருத்துகளை மற்றவர்களுக்குப் புரியும் வழிகளில் வெளிப்படுத்துவதுதான் எழுத்தறிவு

எழுத்தறிவின் செயல்பாடு என்ன? சிந்தனைகள் மற்றும் கருத்துகளை மற்றவர்களுக்குப் புரியும் வழிகளில் வெளிப்படுத்துதல்தானே? படிக்கவும் எழுதவும் தெரியாதவர்கள் சக்தி கட்டமைப்புகளில் பங்கேற்க இயலாது. அவர்களின் குடிமை சார்ந்த குரல்கள் கேட்கப்படுவதில்லை. நிரலாக்கம் தெரியாதவர்களுக்கும், கணினி வழிகளில் சிந்திக்க முடியாதவர்களுக்கும் எதிர் காலத்தில் இந்தக் கதிதானா?

அமெரிக்க அரசாங்கத்தின் முன்னாள் தலைமைத் தொழில்நுட்ப அதிகாரியும் மற்றும் முன்னாள் கூகிள் நிர்வாகியுமான மேகன் ஸ்மித், ஒவ்வொரு குழந்தையும் நிரலாக்கம் செய்யக் கற்றுக் கொள்ள வேண்டுமென்று சொல்கிறார். இந்த உலகத்தைப் புரிந்துகொள்வதற்கும், மாற்றுவதற்கும் தேவை என்பதால் எழுத்துகளையும், எண்களையும் எழுதப் படிக்கக் கற்பிக்கிறோம் என்றால், கணிதம், விஞ்ஞானம், தொழில்நுட்பம் மற்றும் நிரலாக்கம் ஆகியவை அந்த அறிவின் பகுதியாக இருக்க வேண்டும். மாணவர்கள் தாங்கள் தினமும் பயன்படுத்தும் தொழில்நுட்பத்தைப் புரிந்து கொள்ள வேண்டுமெனில், ஆரம்ப கட்டத்திலிருந்தே அந்த கட்டுமானத் தொகுதிகளை அறிமுகம் செய்யத் தொடங்க வேண்டும்.

இளைஞர்களை நுகர்வோருக்கு பதிலாக படைப்பாளர்களாக மாற்ற நிரலாக்கத் திறன் உதவுகிறது. இந்த படைப்புத் திறன் மற்றும் தொழில்நுட்ப எழுத்தறிவு கொண்ட மாணவர்கள் எதிர்காலத்தில் ஆற்றலைத் தக்கவைத்துக்கொள்வர். அவர்கள் அடுத்த தலைமுறை தொழில் முனைவோர் ஆவார்கள்.

எந்தவொரு தயாரிப்பு அல்லது சேவையை வழங்கினாலும் ஒவ்வொரு நிறுவனமும் மென்பொருள் நிறுவனமாக மாறி வருகின்றன

ஒவ்வொரு நிறுவனமும் மென்பொருள் நிறுவனமாக மாறி வருகின்றன. நீங்கள் டயர்கள் அல்லது மின்சாரத்தை உற்பத்தி செய்தாலும் சரி அல்லது உடல்நலன் பற்றிய சேவைகளை வழங்கினாலும் சரி, மென்பொருளை முதன்மையாக வைக்கும் போட்டியாளர்கள் உங்கள் வாடிக்கையாளர்களைத் தங்கள் பக்கம் இழுக்கத் தயாராக உள்ளனர். ஆகவே எந்தவொரு தயாரிப்பு அல்லது சேவை நிறுவனமும் மென்பொருள் நிறுவனமாக மாற வேண்டிய கட்டாயத்தில் உள்ளன.

ஆக ஒவ்வொருவரும் கணினி ஞானம் உள்ளவர்களாக இருக்கவேண்டும் – நுகர்வோராக அல்ல, உருவாக்குநராக.

14. உங்கள் பிள்ளைகளின் கணினி ஐபேடா, ராஸ்ப்பெரி-பையா?

வெறும் பார்வையாளராக மட்டுமே இருக்க விரும்புகிறீர்களா?

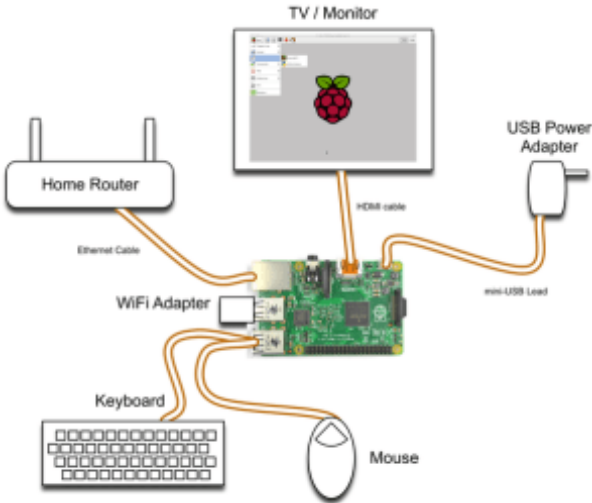
நான் ஏன் ஐபாட் வாங்க மாட்டேன் என்பதற்கு ஒருவர் கூறுகையில், “உங்கள் குழந்தைகளுக்கு ஐபாட் வாங்குதல் என்பது உலகத்தை அக்கக்காகப் பிரித்து தனக்கேற்ற மாதிரி திரும்பவும் முடுக்கிக் கொள்ளலாம் என்பதை உணர்ந்து கொள்வதற்கான ஒரு வழிமுறை அல்ல. அது பேட்டரி மாற்றவேண்டும் என்றால் கூட வேலை தெரிந்த மற்றவர்களை நம்பி இரு என்று உங்கள் பிள்ளைகளுக்கு சொல்லித் தரும் வழி.”

முந்தைய தலைமுறையினரைவிட இளைஞர்கள் அதிக தொழில்நுட்பத்துடன் வளர்ந்து வருகின்றனர் என்பது உண்மை. இருப்பினும், அவர்கள் தொழில்நுட்பத்தை ஈடுபாடற்ற நுகர்வோர்களாக பயன்படுத்துகிறார்களா அல்லது ஆர்வமுள்ள உருவாக்குபவர்களாகவா என்பதுதான் முக்கியம்.

ராஸ்ப்பெரி-பை கணினி ஏன் உருவாக்கப்பட்டது?

2006 ஆம் ஆண்டில் கேம்பிரிட்ஜ் பல்கலைக்கழக கணினி அறிவியல் துறைக்கு விண்ணப்பம் செய்யும் மாணவர்களின் எண்ணிக்கையும் தகுதியும் குறைந்துவிட்டது. கணினி கண்டுபிடிக்கப்பட்ட நாட்டில் (சார்லஸ் பாபேஜ் என்ற ஒரு ஆங்கில இயந்திரப் பொறியியலாளர்தான் நிரலாக்கம் செய்யக்கூடிய கணினி என்ற கருத்தை தொடங்கிவைத்தவர்), கணினி கற்பவர்கள் மிகவும் குறைந்து வருவதாக கல்வியாளர்கள் கவலைப்பட்டார்கள். ஆனால் தேவை என்னவென்றால் கணினி அறிவியல் துறைக்கு மேலும் பலரை ஈர்ப்பது மட்டுமல்ல, கல்லூரியில் சேரும் போது அவர்கள் மேலும் பல பயிற்சிகளில் தேர்ச்சி பெற்றிருக்க வேண்டும்.

ராஸ்ப்பெரி பை அறக்கட்டளையின் நோக்கம் கணினி அறிவியலை குறிப்பாகப் பள்ளிக்கூடங்களில் ஊக்குவிப்பது மற்றும் கணினியை இளையவர்கள் மீண்டும் விளையாட்டாகக் கற்க வழிசெய்வது. கடன் அட்டை அளவிலான இந்தத் தனிநபர் கணினி ஒரு கடினமான பணியை நிறைவேற்ற வேண்டி உள்ளது: இங்கிலாந்து மாணவர்களிடையே கணினி ஆய்வுகளில் ஆர்வத்தை வளர்த்து, சிறுவர்களுக்கு நிரலாக்கம் கற்பித்தல்.



ராஸ்ப்பெரி-பை கணினி

ராஸ்ப்பெரி-பை கணினி எந்த வகையில் வித்தியாசமானது?

இது இங்கிலாந்தில் [ராஸ்பெர்ரி பை அறக்கட்டளையால்](#) உருவாக்கப்பட்ட உள்ளங்கை அளவிலான கணினியின் உள்பாகம் மட்டும். பள்ளிகளிலும் மற்றும் வளரும் நாடுகளிலும் அடிப்படைக் கணினி அறிவியல் கற்பிப்பதை ஊக்குவிப்பதே இவர்களின் குறிக்கோள். 2018 மார்ச் வரை 19 மில்லியன் எண்ணிக்கை விற்று பிரிட்டன் நாட்டுத் தயாரிப்பிலேயே சிறந்த விற்பனையான கணினி ஆகும்.

[தமிழில் ராஸ்பெர்ரி பை பற்றிய அறிமுக காணொளி ஒன்று இங்கே](#). இந்த 3B என்ற புதிய வெளியீட்டில் கம்பியில்லாத் தொடர்பகம் (Wifi) மற்றும் ஊடலை (bluetooth) புதிய அம்சங்கள். இதில் கம்பி மூலம் இணைய இணைப்புக்கு ஈத்தர்வலை (ethernet) இணைப்பானும் உள்ளது. USB இணைப்பான்களைப் பயன்படுத்தி விசைப்பலகை மற்றும் சுட்டியை இணைக்கலாம். HDMI இணைப்பான் மூலம் ஒரு தொலைக்காட்சியை இணைத்து கணினித்திரையாகப் பயன்படுத்தலாம். ஒலி பெருக்கி இணைக்கவும் இணைப்பான் உண்டு. ஆண்ட்ராய்ட் கைபேசிக்கான மின்னேற்றியின் (charger) சக்தியைப் பயன்படுத்தலாம். பொதுநோக்க உள்ளீடு / வெளியீடு (General Purpose Input / Output – GPIO) மின்செருகிகள் 40 உள்ளன. இவற்றின் வழியாக மின்விளக்கு போன்ற வெளிப்புற சாதனங்களைக் கட்டுப்படுத்த முடியும் மற்றும் உணர்கருவிகளையும் இணைக்க முடியும். ஒரு ஐபேட் வாங்கும் செலவில் ஏழெட்டு ராஸ்பெர்ரி பை கணினிகளையும் அவற்றுக்கான துணைக்கருவிகளையும் சேர்த்து வாங்கிவிடலாம்.

பல திறந்த மூல இயங்கு தளங்கள் உள்ளன. ராஸ்பெர்ரி பை அறக்கட்டளை பரிந்துரைப்பது டெபியன் லினக்ஸ் சார்ந்த [ராஸ்பியன் \(Raspbian\)](#). இத்துடன் சேர்ந்து வரும் பைத்தான் நிரலாக்க மொழியையும் பரிந்துரைக்கிறார்கள். ஸ்க்ராட்ச் (Scratch) என்னும் நிரலாக்க மொழி வயதில் குறைந்த சிறுவர்களுக்கு எதிர்காலத்தில் தேவைப்படும் திறன்களை வளர்க்க உதவும். லெகோ (Lego) கட்டுமானத் துண்டுகள் போல அடுக்கிக் கட்டலாம்.

சி-ஸ்டெம் பணியகம் (C-STEM Studio) கணிதத் தல், அறிவியல், தொழில்நுட்பம், பொறியியல் மற்றும் கணிதத்துடன் தானியங்கியியலையும் (robotics) ஒருங்கிணைந்து கற்க கலிஃபோர்னியா பல்கலைக்கழகம் டேவிஸ் உருவாக்கியது, இலவசமாக வழங்கப்படுகிறது. பல இணக்கமான துணைக்கருவிகள், பாகங்கள் இணையத்தில் வாங்க முடியும்

பல திறந்த மூல மென்பொருள் செயலிகள் மற்றும் மென்பொருள் மேம்பாட்டு கருவிகளைப் பதிவிறக்கம் செய்யலாம். [இதன் மன்றத்தில்](#) வெற்றிகரமாக முடிக்கப்பட்ட திட்டங்களைப் பற்றிய தகவலைப் பகிர்ந்துகொள்கிறார்கள் மற்றும் பிரச்சினை வந்தால் கேள்வி கேட்டால் பதில் கிடைக்கும்.

பல நிரலாளர்களும் செயலிகளும் வீட்டைத் தன்னியக்கமாக்கல் (home automation) செய்ய ராஸ்பெர்ரி பையை பயன்படுத்துகிறார்கள். முக்கியமாக மின் நுகர்வைக் கண்காணிக்கவும் குறைக்கவும் செலவு மலிவான தீர்வாக ராஸ்பெர்ரி பையை நிரலாக்கம் செய்கிறார்கள். இது மிகவும் விலையுயர்ந்த வணிக மாற்றுக்கு ஒப்பீட்டளவில் குறைந்த செலவு என்பதால் ஒரு பிரபலமான மற்றும் சிக்கனமான தீர்வாக ஆகியுள்ளது.

வகுப்பறையில் ராஸ்பெர்ரி பை பயன்படுத்த சில வழிகள்

“[ராஸ்பெர்ரி பையில் நிரலாக்கம் கற்றுக்கொள்வது எளிது](#). நான் சமீபத்தில் ஸ்க்ராட்ச் பயன்படுத்தி ஒரு மாணவர்கள் குழுவுக்கு அசைவூட்டமும் மற்றும் பைத்தான் பயன்படுத்தி ஆமை வரைகலை உருவாக்கமும் கற்றுக் கொடுத்தேன். நிரலாக்கம் கற்கத் தொடங்குவதற்குச் சிறந்த வழி பிரைலன் பெய்ன் (Bryson Payne) எழுதிய “[உங்கள் பிள்ளைகளுக்கு நிரலாக்கம் கற்றுக் கொடுங்கள் \(Teach Your Kids to Code\)](#)” என்ற புத்தகம். குறைந்த நேரத்தில் உங்கள் மாணவர்கள் ராஸ்பெர்ரி பையில் பைத்தான் பயன்படுத்தி அழகான வரைகலைகள் செய்யத் தொடங்குவார்கள். இணையத்தை எளிதாக அணுகுவதற்காக ராஸ்பெர்ரி பை 3 குரோமியம் உலாவியுடன் வருகிறது. ராஸ்பியன் பிக்சல் (Pixel) மேசைத்தளம் நிறுவினால் பிரபலமான விளையாட்டு மைன்கிராப்ட் பை (Minecraft Pi), அத்துடன் சேர்ந்து வருகிறது.”

உங்கள் பிள்ளைகள் வெறும் தொழில்நுட்ப நுகர்வோர்களாக மட்டும் இருந்தால் போதுமா அல்லது படைப்பாளர்களாக வேண்டுமா?

உலகம் முழுவதும் மிகப்பெரும்பாலான சிறுவர்கள் தொழில்நுட்பத்தை எவ்வாறு பயன்படுத்துவது என்பதை சிறந்த திறனுடன் கற்றுக் கொண்டிருக்கிறார்கள், ஆனால் தொழில்நுட்பம் எவ்வாறு செயல்படுகிறது என்பதற்கு விழிப்புணர்வு மிகவும் குறைவாகவே உள்ளது. பலருக்கு அவர்கள் பயன்படுத்தும் மற்றும் அனுபவிக்கும் சாதனங்கள் ஒரு மர்மமாகவே உள்ளன. இது உண்மையில் மிகவும் வருத்தமாக உள்ளது. ஏனெனில் இன்றைய மாணவர்கள் தொழில்நுட்ப எதிர்காலத்தில் செழித்து இருக்க வேண்டுமென்றால் அவர்கள் அதன் வெவ்வேறு கோணங்களையும் மற்றும் அவை எவ்வாறு வேலை செய்கின்றன என்பதையும் புரிந்து கொள்ள வேண்டும்.

இதற்கு ஒரு முக்கிய காரணம் கணினியிலும் இணையத்திலும் விளையாடவும், வீடியோவைக் காணவும், இசை கேட்கவும், இணையதளத்தில் குழுவாக ஒத்துழைக்கவும், ஆவணங்கள், படங்கள் மற்றும் திரைப்படங்களைத் திருத்தவும், மற்றும் நீங்கள் எதைச் செய்ய வேண்டுமென்றாலும் ஐபாட் போன்ற அற்புதமான தொழில்நுட்ப சாதனங்கள் சந்தையில் குவிந்து கிடக்கின்றன. இந்த சாதனங்கள் அற்புதமான அனுபவத்தை வழங்கினாலும் இது மிகவும் கட்டுப்படுத்தப்பட்ட அனுபவம். சாதாரண பயனர்கள் சிறப்புக் கருவிகள் இல்லாமல் சாதனங்களைத் திறக்கக் கூட முடியாது. ராஸ்பெர்ரி பை பற்றிய அற்புதமான சங்கதி, எண்ணற்ற பிள்ளைகளுக்கு தங்கள் வானொலியை எவ்வாறு ஒன்று கூட்டுவது மற்றும் மின்னணுவியல் அடிப்படைகளை எப்படிக் கற்றுக்கொள்வது போன்ற கருவிகள் கூட்டுப் பயிற்சிகளை மீண்டும் கொண்டு வருகிறது. ராஸ்பெர்ரி பையின் நிரலாக்கத்தன்மை ஒரு ஊடுருவ இயலாத பளபளப்பான உறை அல்லது சிக்கலான குறியீடுக்குப் பின் மறைக்கப்படவில்லை. பயனர்கள் மாற்றுவதைத் தடுக்கின்ற சட்டங்கள் மூலம் பாதுகாக்கப்பட்ட தனியுரிம மென்பொருட்கள் இதில் ஏற்றப்படவில்லை. மாறாக திறந்த மூல மென்பொருள் மீது அது சார்ந்திருக்கிறது.

நிரலாக்கத்தைக் கட்டாய பாடமாக்கும் முதல் நாடு இங்கிலாந்து ஆகும்

நிரலாக்கத்தைக் கட்டாய பாடமாக ஆக்குவது முக்கியம் என்பதை சில நாடுகள் உணரத் தொடங்கி உள்ளன. இந்த ஆண்டு, 5 வயதிலிருந்து தொடங்கி, 16 வயது வரை தங்கள் பள்ளிகளில் நிரலாக்கத்தைக் கட்டாய பாடமாக்கும் முதல் நாடு இங்கிலாந்து ஆகும்.

லண்டனில் உள்ள [உருவாக்குபவர் பல்கலைக்கழகம்](#) பயிற்சி மூலம் கற்றலையும் ஆசிரியர் பயிற்சியையும் சேர்ந்து அளிக்கிறது. ஒரு வலைப்படக்கருவியை (webcam) உருபெருக்கியாக (microscope) மாற்றுவது எப்படி போன்ற சவால்களில் பயிற்சி வழங்குவதன் மூலம் ஆசிரியர்கள் நடைமுறைக் கூறுகளை கைவேலைத் திட்டங்களில் இணைக்க உதவுகிறது.

15. தகவல் தொழில்நுட்ப உதவியுடன் தமிழை எளிதாகக் கற்பிக்கலாம்

தமிழ் மொழியில் கற்றுக்கொள்ள அதிகமான எழுத்துகள் உள்ளன என்ற ஒரு கருத்து நிலவுகிறது. தமிழில் 12 உயிர், 18 மெய் எழுத்துகள், ஒரு ஆய்த எழுத்து ஆக மொத்தம் 31 எழுத்துகள் உள்ளன. ஒவ்வொரு மெய்யெழுத்தும் ஒவ்வொரு உயிரெழுத்துடன் சேர்ந்து மொத்தம் 216 உயிர்மெய் எழுத்துகளை உருவாக்குகின்றன என்பது உண்மைதான். மெய்யெழுத்தை எழுதி, சேர்க்க வேண்டிய உயிரெழுத்து அடையாளம் 11 தெரிந்தால் போதும். ஆகவே மேற்கண்ட கருத்து சரியா என்று நீங்களே சொல்லுங்கள்.

மொழியைக் கற்பிப்பதில் பல பெரிய மேம்பாடுகளைத் தகவல் தொழில்நுட்பம் கொண்டுவந்துள்ளது. ஆகவே மொழியைக் கற்பிப்பதில் ஏற்கனவே வந்து விட்ட மற்றும் அடுத்து வரும் போக்குகள் என்ன என்று பார்ப்போம்.

திறன் பேசி அல்லது கைக்கணினி உதவியுடன் கற்றல்

நீங்கள் ஒரு மொழியைக் கற்றுக் கொள்ள விரும்பினால், குறுந்தகடுகள் அல்லது புத்தகங்களை விட செயலிகளில் பல வசதிகள் உள்ளன. இவற்றில் மிக முக்கியமானவை அவற்றின் ஊடாடல் (interactive) பண்புகள். எங்கு சென்றாலும் இவை உங்கள் கையிலிருப்பதால் பயணம் செய்யும் போதும், வரிசையில் காத்திருக்கும் போதும் நேரத்தை வீணாக்காமல் பயன்படுத்தலாம்.

திறன் பேசி அல்லது கைக்கணினிகளில் கற்பதை ஆங்கிலத்தில் MALL (Mobile Assisted Language Learning) என்று சொல்கிறார்கள். இதில் பல வசதிகள் உள்ளன. நீங்கள் கற்றுக் கொள்ளும் மொழியில் ஒரு உரையாடலுக்குச் சற்று முன்னர் சட்டென்று ஒரு சிறு பயிற்சி செய்து கொள்ளலாம். உரையாடல் முடிந்தவுடன் எந்த இடத்தில் சிறு முன்னேற்றங்கள் செய்யலாம் என்று பயிற்சி செய்யலாம். மேலும், ஒத்திசைந்த குழுக்களில் கற்றல் செயல்முறையைப் பகிர்ந்து கொள்வது எளிது.

எல்லா மாணவர்களும் ஒரே மாதிரியாகக் கற்றுக் கொள்வதில்லை, ஒரே வேகத்தில் கற்றுக் கொள்வதில்லை. எனவே, மொழி கற்றல் பயிற்சிகள், ஒவ்வொரு மாணவரின் முன்னேற்றத்திற்கும் ஏற்படையதாக இருக்க வேண்டும். பலவீனமான இடங்களை வலுவூட்ட வேண்டும், ஏற்கனவே தெரிந்த அல்லது விரைவாகக் கற்றுக் கொண்ட தலைப்புகளைக் கடந்து செல்லலாம். ஒருவருக்குக் கடினமாக இருக்கும் அதே இடத்தில் இன்னொருவருக்கு இருப்பதில்லை. ஆகவே கற்பதற்கான வழிமுறைகளைப் பயனருக்குத் தகுந்தவாறு தனிப்பயனாக்கலாம்.

கற்றலை விளையாட்டாக ஆக்குதல்

கற்றலை விளையாட்டாக ஆக்கி மதிப்பெண் கொடுத்தால் சக மாணவர்களுடன் போட்டி போட்டுக் கொண்டு கற்கவும் ஈடுபாடு கொள்ளவும் வாய்ப்பு அதிகம். விளையாட்டாக்குதலைத் திறம்படச் செய்ய நிறைய படைப்பாற்றலும் கற்பனையும் தேவை.

டுவோலிங்கோ (Duolingo) மற்றும் மைண்ட்ஸ்நாக்ஸ் (MindSnacks) போன்ற இயங்குதளங்கள் கற்பவர்களை ஈடுபடுத்துவதற்கும், அவர்கள் மொழியைத் திறம்படக் கற்பதற்கும் விளையாட்டாக்குதலைப் பயன்படுத்துகிறார்கள். ஈடுபாட்டு நிலைகளை விளையாட்டாக்குதல் அதிகரிக்கிறது மற்றும் “வேடிக்கைத்” தன்மை தொடர்ந்து இருப்பதால் மாணவர்கள் ஆர்வத்துடன் கற்றுக் கொள்கிறார்கள்.

தொடர்ந்து பின்னூட்டங்கள் கொடுத்தல்

அடிக்கடி பின்னூட்டங்கள் கொடுப்பது மாணவர்களை சரியாக வழிநடத்தவும் தேவையான மேம்பாடுகளைச் செய்யவும் உதவுகிறது. மாணவர்கள் எவ்வாறு செயல்படுகிறார்கள் என்பதைப் புரிந்துகொள்வது மட்டுமல்லாமல், பயிற்றுநர்கள் சிறந்த வழிமுறையை அடைவதற்கும் உதவும்.

பல்வேறு இணையக் கருவிகள் கல்வியாளர்களை நேரடியாக தனிப்பட்ட கருத்துகளை வழங்க அனுமதிக்கின்றன, இதனால் மாணவர்கள் அதிக கவனம் செலுத்த வேண்டிய இடங்களில் எளிதாக உடன் அடையாளம் காட்ட முடியும்.



மொழிப்பயிற்சி

மொழிப் பயிற்சிக்கு டுவோலிங்கோவும் (Duolingo) மெம்ரைசும் (Memrise) மைண்ட்ஸ்நாக்சும் (Mindsnacks)

[டுவோலிங்கோவில் \(Duolingo\) மைண்ட்ஸ்நாக்சிலும் \(Mindsnacks\)](#) இன்னும் தமிழ்ப் பயிற்சி வரவில்லை. [மெம்ரைசில் ஓரளவுதான்](#). எனினும் டுவோலிங்கோவில் 500 மில்லியன் பேருக்கு மேல் பல மொழிகளைக் கற்றுக் கொண்டிருக்கிறார்கள். எனவே நவீன தொழில்நுட்பம் பற்றி அறிய இவையே தலைசிறந்தவை. டுவோலிங்கோ திறந்த மூலம் அல்ல. ஆனால் மொழியை இலவசமாகக் கற்றுக் கொள்ளலாம். ஆன்டிராய்டு, ஐபோன், ஐபாட் செயலி அல்லது இணைய தளத்தில் கற்றுக் கொள்ளலாம். ஒவ்வொரு பாடத்தின் இறுதியிலும் விளம்பரங்கள் வரும். கட்டணம் கட்டினால் விளம்பரங்கள் வராது. தன்னார்வலர்கள் பங்களிப்பை வைத்து, இந்தி பேசுபவர்கள் ஆங்கிலம் பயில்வதும் சேர்த்து, இதுவரை 73 மொழி இணைகள் வந்து விட்டன. தமிழ் பேசுபவர்கள் ஆங்கிலம் பயில் வேலை நடந்து கொண்டிருக்கிறது. ஆனால் தமிழ் பயிற்சிக்கு வேலை தொடங்கவில்லை.

உள்நுழையும்போது நீங்கள் என்ன கற்றுக் கொள்ளப் போகிறீர்கள் என்பதற்கான ஒரு வரைபடத்தை முகப்புத் திரையில் காட்டுகிறது. அடுத்து உங்களுடைய இலக்கை உள்ளிடும்படி அறிவுறுத்துகிறது. எடுத்துக்காட்டாக ஒவ்வொரு நாளும் கற்பது, நாளொன்றுக்கு 30 புள்ளிகள் சம்பாதிப்பது. ஒரு பாடத்துக்கு பத்து புள்ளிகள் சுமார் ஐந்து நிமிடங்கள் எடுக்கும். உங்கள் இலக்குகளை நீங்கள் அடைந்தால், திரையில் மகிழ்ச்சியான செய்திகளை அளிப்பதுடன் கூடுதல் உள்ளடக்கங்களையோ அல்லது அம்சங்களையோ அணுக வழி செய்கிறது.

பயிற்சிகள் ஒன்றன் மேல் ஒன்றாகக் கட்டப்பட்டுள்ளன. உங்கள் முன்னேற்றம் படி சரியான நேரத்தில் சொற்கள் மற்றும் கருத்தாக்கங்கள் அறிமுகப்படுத்தப்படும். பன்முகத் தெரிவு வினாக்கள், ஒலியைக் கேட்டு எழுதும் பயிற்சிகள், மொழிபெயர்ப்பு பயிற்சிகள் மற்றும் பேச்சுப் பயிற்சிகள் ஆகியவையும் உள்ளன. இம்மாதிரிப் பயிற்சி புதிதாக ஒரு மொழியைக் கற்றுக் கொள்ளத் தொடங்குபவர்களுக்கு மிகவும் பொருத்தமானது

மெம்ரைசு ஒரு சொல்லை அதன் பொருளுடன் இணைக்கப் பல உத்திகளைக் கையாளுகிறது. ஒரு பொதுவான எடுத்துக்காட்டு புதிய சொற்களை ஞாபகப்படுத்த உதவுவதற்கு நினைவூட்டல்கள்.

தமிழ் கற்க மற்ற செயலிகளும் இணைய தளங்களும்

[தமிழை சீக்கிரம் கற்றுக்கொள்ளுங்கள் \(Learn Tamil Quickly\)](#) சுமார் அரை மில்லியன் பேர் பதிவிறக்கி 5000 பேர் தரமதிப்பீடு செய்து 5 க்கு 4.3 வாங்கி முதலிடத்தில் உள்ளது “தமிழை சீக்கிரம் கற்றுக்கொள்ளுங்கள் Learn Tamil Quickly” செயலி. ஆன்டிராய்டு மற்றும் ஐபோன் / ஐபாட் ஆகியவற்றின் செயலிகளை இலவசமாக பதிவிறக்கம் செய்யலாம். ஆனால் பாடங்களில் விளம்பரங்கள் உண்டு. வினாடி-

வினாக்கள், நினைவூட்ட அட்டைகள் (flash cards) மற்றும் உங்கள் சொந்த சொற்றொடர்களை பதிவு செய்வதன் மூலம் கற்கலாம். உடல், நேரம், கால அட்டவணை, நிறங்கள், திசைகள், உடல்நலம் போன்ற 33 பிரிவுகளில் 1200 க்கும் மேல் தமிழ்ச் சொற்கள் மற்றும் சொற்றொடர்கள். தமிழைத் தாய்மொழியாகப் பேசுபவர்களின் தெளிவான உச்சரிப்பின் தரமான ஒலிப்பதிவு உங்கள் உச்சரிப்பு மற்றும் நேர்த்தியான குரல் ஒலிக்கும் உதவும். மற்றும் [கேலு \(Kelu\)](#) செயலியை சுமார் நூறாயிரம் பேர் பதிவிறக்கி 800 பேர் தரமதிப்பீடு செய்துள்ளனர்.

[“தமிழ் 2.0 தமிழ் எழுத்துகள்” ஆன்டிராய்டு செயலி](#) நினைவூட்ட அட்டைகள் மூலம் தமிழ் எழுத்துகளைக் கற்றுக் கொள்ள உதவும். இது சிறுவர்களுக்குப் பிடித்த மாதிரி அசைவூட்டம் மற்றும் ஊடாடல் கொண்டது.

இந்தப் பாடத்திட்டங்கள் மூலம் ஓரளவு முன்னேற்றம் அடைந்தபின், உங்கள் சொற்களஞ்சியம் அதிகரித்து, ஏற்கனவே நீங்கள் கற்றுக்கொண்டதைத் தக்கவைத்துக் கொள்ள வேண்டியது முக்கியம் என்ற ஒரு நிலையை நீங்கள் அடைவீர்கள். இதற்கு மனப்பாடம் செய்ய உதவும் [திறந்த மூல செயலி ஆன்கி](#) ஏற்றது. மற்றவர்களால் உருவாக்கப்பட்ட தமிழ்ச் சொல் பட்டியலைப் பதிவிறக்கி உங்களுக்குப் பிடித்தவாறு மாற்றலாம், அல்லது உங்கள் சொந்தச் சொல் பட்டியலை உருவாக்கலாம். இணையதளம் மற்றும் ஆன்ட்ராய்டு, ஐபோன் செயலிகள் உள்ளன.

நீங்கள் ஒரு ஆசிரியருடன் நேரடியாகவோ அல்லது தொலைபேசியிலோ இருந்தால், அவர்கள் உங்கள் உச்சரிப்பைக் கேட்டுத் திருத்த முடியும். அது இல்லாவிட்டால், கூகிள் ஜிபோர்ட்டில் (GBoard) தமிழ் சொல்வதெழுதல் (dictation) திறன் உள்ளது. இதை ஆன்ட்ராய்டு, ஐபோன், ஐபாடில் நிறுவலாம். இதில் உங்கள் உச்சரிப்பை முயற்சி செய்து, சரியான சொற்களை அடையாளம் காண்பதை உறுதி செய்யலாம்.

நீங்கள் எப்படியும் இசையைக் கேட்கிறீர்கள், அதேபோல் திரைப்படம் மற்றும் தொலைக்காட்சியையும் பார்க்கிறீர்கள். ஓரளவு பரிச்சயமானபின், இதை ஏன் கற்றுக்கொள்ளும் புது மொழியில் செய்யக்கூடாது? நீங்கள் முதலில் புரிந்து கொள்ளப் போவதில்லை, ஆனால் இந்தக் கட்டத்தில் அது அவ்வளவு முக்கியம் இல்லை. இங்கே நாம் செய்ய முயற்சிப்பது ஒரு புதிய மொழியின் நுணுக்கங்களில் காதுகளுக்குப் பயிற்சி கொடுப்பதுதான்.

16. தகவல் தொழில்நுட்ப உதவியுடன் நீங்களும் ஒரு எழுத்தாளராகலாம்

வெளியீடு செய்த எழுத்தாளராக ஆவதற்கு இதுதான் வரலாற்றிலேயே சிறந்த காலம் என்று நீங்கள் கேள்விப்பட்டிருக்கலாம். அது உண்மைதான். நூலாசிரியர்கள் முன்னர் இருந்ததை விட வாசகர்களை அடைய அதிக வாய்ப்புகளை கொண்டுள்ளனர். மேலும் தங்கள் படைப்புகளை வெளியீடு செய்வதில் முன்னை விட அதிகமான கட்டுப்பாட்டை எடுத்துக் கொள்ள விரும்பினால், அதுவும் சாத்தியமே.

மேலும் புத்தகங்களை விநியோகம் செய்வதில் வந்த மாற்றங்களால் ஒவ்வொரு வாசகருக்கும் எந்தப் புத்தகமும் எங்கு வேண்டுமானாலும் எப்போது வேண்டுமானாலும் கிடைக்கும். இணையப் புத்தகக் கடை அலமாரிகள் எல்லையற்றதாகவும் மற்றும் உலகளாவியதாகவும் உள்ளன. தேவைப்படும் போது அச்சிடும் (print on demand) தொழில்நுட்பம் ஒரு புத்தகம் எப்பொழுதுமே அச்சில் இருக்க உதவுகிறது.

உங்களுக்கு இயல்பான படைப்பாற்றல் இருந்தால் சிறிய புனைகதைகளும் முழுநீளப் புதினங்களும் எழுதலாம். “ஒரு புத்தகத்தை நீங்களே வெளியிடுவது எப்படி” என்பது போன்ற உதவிக் குறிப்புகள் இணையதளத்தில் பல உள்ளன. ஒருக்கால் கடினமாக இருந்தால் மனச்சோர்வு அடைய வேண்டாம். முதலில் சிறிய அளவில் தொடங்குங்கள்.

விக்கிப்பீடியா கட்டுரைகளை எழுதி அல்லது தமிழாக்கம் செய்து பழகிக் கொள்ளுங்கள்

முழுப் புத்தகம் எழுதுவது பற்றி நினைத்தாலே மலைப்பாக இருக்கிறதா? உங்களுக்கு மிகவும் நன்றாகத் தெரிந்த தலைப்பில் விக்கிப்பீடியாவில் கட்டுரைகள் எழுத ஆரம்பியுங்கள். ஆங்கில விக்கிப்பீடியாவில், அல்லது உங்களுக்கு நன்றாகத் தெரிந்த வேறு மொழியில், உள்ள கட்டுரைகளைத் தமிழாக்கம் செய்யுங்கள். தமிழ் விக்கிப்பீடியா நிர்வாகிகளின் பின்னூட்டத்தை பயன்படுத்தி உங்கள் எழுத்துத் திறமையை நன்கு வளர்த்துக் கொள்ள முடியும். ஆங்கிலச் சொல்லுக்கு ஒப்பான தமிழ்ச் சொல் கிடைப்பது கடினமாக இருக்கிறதா? [தமிழ் விக்சனரியில் தேடுங்கள். தமிழ் விக்சனரியில் 350 ஆயிரத்துக்கு மேல் பக்கங்கள் உள்ளன.](#)



எழுத்தாளர்

கட்டற்ற திறந்த மூல மென்பொருட்களின் பயனர் இடைமுகம் மற்றும் கையேடுகள் எழுதலாம் அல்லது தமிழாக்கம் செய்யலாம்

நீங்கள் தொழில்நுட்ப ஆர்வலரா? கட்டற்ற திறந்த மூல மென்பொருட்களின் பயனர் இடைமுகத்தை மொழிபெயர்க்கவும் கையேடுகளை எழுதவும் மற்றும் மொழிபெயர்க்கவும் வாய்ப்புகள் எப்பொழுதும் உண்டு. திறந்த மூல மென்பொருட்களின் தன்னார்வலர் மடலாடற் குழுக்களில் சேருங்கள். உங்களுக்குப் பிடித்த வேலையைச் செய்யலாம். அதே நேரத்தில் புதிய தொழில்நுட்பங்களைக் கற்றுக் கொள்ளலாம். புதிய நண்பர்களும் வழிகாட்டிகளும் கிடைப்பார்கள்.

உங்களுக்கு மிகவும் பிடித்த நூலைத் தமிழாக்கம் செய்து வெளியிடுங்கள்

பல மொழிகளிலிருந்து முக்கிய இலக்கியப் படைப்புகளை தமிழாக்கம் செய்யும் பாரம்பரியம் சில பத்தாண்டுகளுக்கு முன் நம் சமூகத்தில் இருந்துள்ளது. ஆனால் சமீபத்தில் ஏனோ ஆர்வமிழந்துள்ளது. “லார்ட் லிட்டன் அவர்களால் எழுதப்பட்ட ‘தி சீக்ரட் வே’ என்ற நூலை ‘மனோன்மனீயம்’ என்ற நாடகமாக எழுதினார் பேராசிரியர் சுந்தரம். சுமார் அறுபது ஆண்டுகளுக்கும் முன்பாகவே பிரெஞ்சு *Les Miserables* என்ற அற்புதமான நாவல் தமிழில் வெளிவந்தது. காங்கிரஸ் தலைவர்களில் ஒருவரான ஓ.வி. அழகேசன் மொழிபெயர்த்தார். குமாரசாமி, சேனாபதி ஆகியோர் வங்க இலக்கியங்களைத் தமிழில் தந்தனர். கா.மூ.மூ. மராட்டிய மொழியிலிருந்து தமிழுக்கு மொழிபெயர்த்தார். காண்டேகரின் படைப்புகள் அனைத்தையும் தமிழுக்குக் கொண்டு வந்தார். ரா. வீழிநாதன், சரசுவதி ராம்நாத் ஆகியோர் இந்தி மொழியிலிருந்து தமிழுக்கு பல படைப்புகளைக் கொண்டுவந்தனர்.”

மொழிபெயர்ப்பாளருக்கான சாகித்திய அகாதெமி விருது பெற்ற க. பூரணச்சந்திரன்

“நீட்சே மிகச்சுருக்கமான அறிமுகம்”, “காந்தியைக் கொன்றவர்கள்”, “நொறுங்கிய குடியரசு” போன்ற பல நூல்களைத் தமிழாக்கம் செய்து “பொறுப்புமிக்க மனிதர்கள் (*Serious Men*)” என்ற நூலுக்கு 2016 ஆம் ஆண்டிற்கான தமிழின் சிறந்த மொழிபெயர்ப்பாளருக்கான சாகித்திய அகாதெமி விருது பெற்ற க. பூரணச்சந்திரன் அவர்கள் தனது மொழிபெயர்ப்பு அனுபவங்களை மிக விரிவாக இங்கே பகிர்ந்து கொண்டுள்ளார். இதோ சில எடுத்துக்காட்டுகள். “இன்றைக்கு மொழிபெயர்ப்பின் தேவையையும் இன்றியமையாமையையும் யாரும் எடுத்துச் சொல்லத் தேவையில்லை. செய்தித்தாள்கள், தொலைக்காட்சி முதலாகப் பாடப்புத்தகங்கள் வரை நமக்குக் கிடைக்கும் தகவல்களில் பெரும் பகுதி மொழிபெயர்ப்பின் வாயிலாகத்தான் கிடைக்கின்றன.” “வாசகரைவிட, மொழிபெயர்ப்பாளருக்கு ஒத்துணர்வு (*empathy*) மிகவும் கூடுதலாகத் தேவை. ஒத்துணர்வு இல்லாவிட்டால் ஆசிரியரைப் புரிந்து கொண்டு மொழிபெயர்க்க இயலாது.”

“பெயரெச்சம், வினையெச்சம், வினையாலணையும் பெயர் என்ற மூன்றும் தமிழில் மிகச் சுருக்கமாகச் செய்திகளைச் சொல்ல வல்லவை. *The men who killed Gandhi* என்ற தொடரை நான் காந்தியைக் கொன்றவர்கள் என்று சுருக்கமாக முடித்து விட்டேன். *The men who killed* என்ற நீண்ட தொடருக்கு, தமிழில் ஒரே ஒரு வினையாலணையும் பெயர் – கொன்றவர்கள் என்பது போதுமானதாகி விடுகிறது. *The man who came yesterday* என்பதை நேற்று வந்தவர் என்று இரண்டே சொற்களில் அடக்கலாம். (பெயரெச்சம்) *She ran and fell down* என்பதை ஓடி விழுந்தாள் என்று சொல்லலாம். ஓடி என்பது வினையெச்சம். *To cut a long story short* என்பதை வளர்த்துவானேன் என்பாராம் புதுமைப்பித்தன். ஆங்கிலத்தில் ‘ரிலடிவ் க்ளாஸ்’ எனப்படுவனவற்றைத் தமிழில் மிகச் சுருக்கமாக மொழிபெயர்க்க இம்மூன்றும் உதவுகின்றன... இவை மூன்றையும் சரிவரக் கையாளத் தெரியாதவர்கள்தான் ஆங்கிலத்தில் சுருக்கமாகச் சொல்லமுடிகிறது, தமிழில் சொன்னால் அது நீளமாகிறது, அதிகரிக்கிறது என்பார்கள்.”

பதிப்புரிமையும் பொதுவுரிமையும்

ஒரு நூலின் பதிப்புரிமை காலம் முடிவடைந்தால் அது பொதுவுரிமையில் வந்து விடும். அதை எவரும் மொழிபெயர்த்துப் பதிப்பிக்கலாம். விக்கிப்பீடியாவில் பல்வேறு நாடுகளுக்கான பதிப்புரிமை ஆண்டுகளும் விதிமுறைகளும் கொண்ட பட்டியல் உள்ளது. பொதுவுரிமையில் வந்து விட்ட ஆங்கிலப் புத்தகங்களின் பட்டியல் இங்கு கிடைக்கிறது. ஏன் அவ்வளவு தூரம் போக வேண்டும்? இந்த இணைய தளத்தில் பொதுவுரிமையில் உள்ள ஆங்கில நூல்களைப் பதிவிறக்கமே செய்யலாம். இவற்றில் மிகவும் பிரபலமான பல ஏற்கனவே தமிழாக்கம் செய்யப்பட்டு விட்டன. எனினும் மேலும் தமிழாக்கம் செய்ய வாய்ப்புகளும் பல உள்ளன. நீங்கள் மொழிபெயர்த்து வெளியிட மிக விருப்பப்படும் நூல் பதிப்புரிமையில் உள்ளது என்றால் என்ன செய்வது? அதற்கான உரிமைகளைப் பெறுவது எப்படி என்ற கையேட்டை இங்கே பார்க்கலாம்.

எடுத்துக்காட்டாக தமிழாக்கம் செய்து மின்னூலாக சமீபத்தில் freetamilebooks.com வெளியிட்ட நூல் ஒன்று இதோ. சுவர்கத்தின் நுழைவாயில் – தமிழில்: சே.அருணாசலம் (ஆங்கில முதன்நூல்: *ENTERING THE KINGDOM* by James Allen (1903). நீங்கள் எழுதிய உரையை மின்னூலாக்குவது எப்படி என்று இந்த

இணையதளத்தில் படிப்படியாக நிகழ்படம் பார்த்து நீங்களே செய்யலாம். ஆன்டிராய்டு, ஐபோன், ஐபாட், கிண்டில் போன்ற கைச் சாதனங்களில் படிக்கக்கூடிய எல்லா வடிவங்களிலும் வெளியீடு கிடைக்கும். குறைந்த செலவில் ஒருசில பிரதிகள் மட்டும் அச்சிட்டு புத்தகமாகத் தரவும் சேவைகள் உள்ளன.

கணினி உதவி மொழிபெயர்ப்பு

[ஒமேகா-டி \(OmegaT\) என்பது ஒரு கட்டற்ற மற்றும் திறந்த மூல கணினி உதவி மொழிபெயர்ப்புத் \(Computer Assisted Translation\) கருவி ஆகும்.](#) இது விண்டோஸ், மேக் அல்லது லினக்ஸ் கணினிகளிலும் இயங்கும்.

ஒரு மொழிபெயர்ப்பில் அதே களத்தைச் சேர்ந்த மற்ற மொழிபெயர்ப்புத் திட்டங்களில் உருவாக்கிய சொற்களஞ்சியங்களைப் பயன்படுத்தலாம். இது குழு மொழிபெயர்ப்பை சீர்மைப்படுத்த உதவுகிறது. ஒருவர் மொழிபெயர்ப்பு செய்து இன்னொருவர் சரிபார்க்க வேண்டுமானால் இது மிகவும் வசதியானது. ஒவ்வொரு வாக்கியமாக எடுத்து அதன் மூலத்தையும் மொழிபெயர்த்த தமிழ் வாக்கியத்தையும் ஒன்றன் கீழ் ஒன்றாகக் காட்டும். இது சரிபார்க்கவும் திருத்தங்கள் செய்யவும் வசதியானது. மைக்ரோசாஃப்ட், கூகிள் (பணம் செலுத்திய கணக்கு), மற்ற சில இயந்திர மொழிபெயர்ப்புக் கணக்கை உள்ளிட்டால் தானியங்கியாக இயந்திர மொழிபெயர்ப்பைப் பெற்றுத் தரும். மைக்ரோசாஃப்ட் வேர்ட், எக்செல், பவர்பாயிண்ட், மீஉரைக் குறிமொழி (HTML), லிபர் ஆபீஸ், ஒபன் ஆபீஸ், விக்கிப்பீடியா கட்டுரைகள், வடிவமைக்கப்படாத உரை போன்ற 30 க்கும் மேற்பட்ட கோப்பு வடிவங்களுடன் வேலை செய்கிறது.

17. உரையும் பேச்சும் கொண்ட மொழித்தொகுப்பு

மொழியியல் பகுப்பாய்வுக்கு அடிப்படையாக இருக்கும், உரையும் பதிவு செய்த பேச்சும் கொண்ட தொகுப்புகளை, மொழித்தொகுப்பு (*corpus*) என்று சொல்கிறோம்.

ஆங்கில மொழித்தொகுப்புகளின் வரலாறு

100 மில்லியன் சொற்கள் கொண்ட பிரிட்டானிய நாட்டு மொழித்தொகுப்பு (BNC), பர்மிங்ஹாம் மொழித்தொகுப்பு, லன்காஸ்டர் ஆங்கில பேச்சுத் தொகுப்பு முதலிய தொகுப்புகள் ஆங்கில மொழிக்குப் பிரபலமானவை. இருமொழி மொழித்தொகுப்புகள் இரண்டு மொழிகளின் மொழிபெயர்ப்பைக் கொண்டிருக்கும். ஐரோப்பிய ஒன்றிய முன்முயற்சி (ECI) பலமொழித்தொகுப்பு ஆகும். இது துருக்கிய, ஜப்பானிய, ரஷ்ய, சீன மற்றும் பிற மொழிகளில் 98 மில்லியன் சொற்களைக் கொண்டுள்ளது.

COBUILD ஆங்கில மொழித்தொகுப்பு

COBUILD என்பது 1980 ஆம் ஆண்டில், காலின்ஸ் பதிப்பு நிறுவனமும் பர்மிங்காம் பல்கலைக்கழகமும் இணைந்து நிறுவிய ஆராய்ச்சி மையமாகும். இத்திட்டத்தின் மிக முக்கியமான சாதனை சமகால உரைகளின் மின்னணுத் தொகுப்பான காலின்ஸ் உரைத்தொகுப்பு உருவாக்கம் மற்றும் பகுப்பாய்வு ஆகும். பின்னர் இது ஆங்கில வங்கி உரைத்தொகுப்பு வளர்ச்சிக்கும் வழிவகுத்தது. பல அகராதிகள் மற்றும் இலக்கண நூல்களும் வெளியிடப்பட்டுள்ளன. இவை அனைத்துமே ஆங்கில வங்கி உரைத்தொகுப்பு ஆதாரங்களை அடிப்படையாகக் கொண்டவை.

இந்த COBUILD உரைத்தொகுப்பு 4.5 பில்லியன் சொற்கள் கொண்ட ஆங்கில உரைகளின் தொகுப்பு. இவை பெரும்பாலும் பிரிட்டானிய உரைகள். ஆனால் வட அமெரிக்கா, ஆஸ்திரேலியா, நியூசிலாந்து, தென்னாப்பிரிக்கா மற்றும் பிற காமன்வெல்த் நாடுகளிலிருந்தும் உள்ளடக்கம் சேர்க்கப்பட்டுள்ளது. பெரும்பாலான உரைகள் வலைத்தளங்கள், பத்திரிகைகள், இதழ்கள் மற்றும் புத்தகங்கள் ஆகியவற்றிலிருந்து சேகரிக்கப்பட்ட ஆங்கில மொழியில் எழுதப்பட்ட உரைகளாகும். ஆனால் வானொலி, தொலைக்காட்சி மற்றும் இயல்பான உரையாடல்களிலிருந்தும் பேச்சுகள் பதிவு செய்யப்பட்டு உள்ளன. ஆங்கில வங்கியில் மொத்தம் 650 மில்லியன் சொற்கள் உள்ளன. பதிப்புகள் ஹார்பர் காலின்ஸ் வெளியீட்டாளர்கள் மற்றும் பர்மிங்காம் பல்கலைக்கழகம் ஆகிய இரண்டிலும் உள்ளன. பர்மிங்காமில் உள்ள பதிப்பை கல்வி சார்ந்த ஆராய்ச்சிக்கு அணுகலாம்.

பிரிகேம் யங் பல்கலைக்கழகத்தில் மொழியியல் பேராசிரியரான மார்க் டேவிஸ் சமகால அமெரிக்க ஆங்கில மொழித்தொகுப்பை (COCA) உருவாக்கினார். இது 160,000 க்கும் அதிகமான உரைகளிலிருந்து சுமார் 560 மில்லியன் சொற்களை உள்ளடக்கியது. 1990 முதல் ஒவ்வொரு ஆண்டும் இதில் 20 மில்லியன் சொற்களைச் சேர்க்கிறார்கள். ஒவ்வொரு மாதமும் இதைப் பல்லாயிரக்கணக்கானோர் பயன்படுத்துகின்றனர். எனவே தற்போது மிகவும் பரவலாக பயன்படுத்தப்படும் மொழித்தொகுப்பு என்று இதைச் சொல்லலாம்.

85 மில்லியன் சொற்கள் பேச்சு, 81 மில்லியன் சொற்கள் புனைகதை, 86 மில்லியன் சொற்கள் பிரபலமான இதழ்களிலிருந்து, 81 மில்லியன் சொற்கள் செய்தித்தாள்களிலிருந்து. 81 மில்லியன் சொற்கள் கல்வி சார்ந்தப் பத்திரிகைகளிலிருந்து. அமெரிக்க காங்கிரஸ் நூலகம் உரைகளை சட்டம், கல்வி, அறிவியல், தொழில்நுட்பம், மருத்துவம், வேளாண்மை, இசை போன்ற வகைப்பாடு செய்கிறது. இந்த முறையின் முழு வரம்பையும் உள்ளடக்குமாறு இந்த உரைகள் தேர்ந்தெடுக்கப்பட்டன.

text_id	Left context	Keyword	Right context
Kalki_Sivakamiyin_Sat	முடிய வானடியாக வந்த இராஜ குலத்தின் வீரக் கணையோடு சிறந்த	கல்வி	ஆராய்ச்சியினாலும் கலைப் பயிற்சியினாலும் ஏற்படும் வித்யாதேஜனும் பிரகாச
Kalki_Sivakamiyin_Sat	றக்கு நூற்றைம்பது வருஷங்களுக்கு முன்னாம் உன்னைப் போலவே	கல்வி	பயிர்வதற்காக, மயூரசன்மன் என்னும் இளைஞன் இந்த நகருக்கு வந்தான். அவ
Kalki_Thyaga_Bhoomi	வந்தன. ஒரு நகரத்தில் இவ்வளவு அநாதாகிரமங்களும், இவ்வசக்	கல்வி	ஸ்தாபனங்களும், மாதர் முன்னேற்றக் கழகங்களும், அமெச்சூர் நாடக சபைகள்
Mu_Va_Akal_Vilakku	முதல் பஸ் விடுபவன் வரையில், பள்ளிக்கூட ஆசிரியர் முதல்	கல்வி	அதிகாரி வரையில் சத்திரஸிடமே மிக்க அன்பு செலுத்தியதைக் கண்டேன். வந்த
Mu_Va_Akal_Vilakku	. " " உயர்ந்த கருத்துத்தான். ஆனால் அப்படி	கல்வி	ஒன்றையே நாடி அடங்கியிருந்தால் பொது அறிவு வளராமல் போகுமே. " " கற்று
Mu_Va_Akal_Vilakku	. செல்வம் இன்றைக்கு இருக்கும்; நாளைக்குப் போகும்.	கல்வி	அப்படிப்பட்டது அல்ல. யாரோ ஒருவரைத்தான் கலைமகள் தேடி வருவாள். உங்
Mu_Va_Akal_Vilakku	வேறு எதிலும் ஈடுபடக்கூடாது " என்றார். " இந்தக்	கல்வி	எல்லாம் சொல்லி வரக்கூடாது. சொல்லாமலே பார்த்துக் கற்றுக் கொள்ள வேண்டு
Mu_Va_Akal_Vilakku	க்கத்து ஊர்களிலும் உயர்நிலைப்பள்ளிகள் ஏற்பட்டு விட்டபடியால்	கல்வி	காரணமாக வருவோரும் குறைத்து விட்டார்கள். முன் இந்தக் குறை எல்லாம் இவ்
Mu_Va_Akal_Vilakku	இவ்வளவு ஏன்? சத்திரனை என் தாய் அடிக்கடி பாராட்டியதற்கும்	கல்வி	அதிகாரி முதல் நெல்லிக்காய் விற்பவர் வரையில் எல்லோரும் போற்றியதற்கும்
Mu_Va_Akal_Vilakku	நெயர்கள், அவர்களை ஒழுங்குபடுத்த முடியாமல் துன்புறுகிறார்கள்.	கல்வி	நிலையங்களில் உரிமையைப் போற்றினால், கல்வி தவிர மற்றத் துறைகளில் ஊக்
Mu_Va_Akal_Vilakku	துன்புறுகிறார்கள். கல்வி நிலையங்களில் உரிமையைப் போற்றினால்,	கல்வி	தவிர மற்றத் துறைகளில் ஊக்கம் மிகுந்து உண்மைப் பயன் குறைந்து மானவர்க
Mu_Va_Akal_Vilakku	மாடப் பார்த்துக் கொடுக்க முயற்சி செய்கிறவர்கள் என்றும், குணம்	கல்வி	முதலியவைகளைப் பெற்றோர்கள் பாப்பதில்லை என்றும் நீ அன்று ஒருநாள் செ
Nonfiction_articles	ந்திய ஆட்சியைச் சாரா. சென்னைப் பல்கலைக் கழகத்தில் கல்லூரிக்	கல்வி	நிலையில் தமிழுக்கு முதன் முதலாக இடம் தேடித் தந்தவரும், சென்னை நகரில்

தமிழ் உரைத்தொகுப்பு

தமிழ் மொழித்தொகுப்புகள்

மைசூர் மத்திய இந்திய மொழிகள் கழகத்தில் 3 மில்லியன் சொற்கள் கொண்ட தமிழ் உரைத்தொகுப்பு ஒன்று கிடைக்கிறது. கல்வியாளர்களுக்கும் கல்வி நிறுவனங்களுக்கும் இலவசமாக விநியோகிக்கப்படுகிறது. பதிவிறக்கம் கிடையாது, எழுதிக் கேட்க வேண்டும். இருப்பினும், எதிர்காலத்தில் இந்நிறுவனம் இந்திய மொழிகளில் கட்டணம் வாங்கிப் பல்வேறு வகையான வளங்களை வழங்க முடியும் என்று சொல்கிறார்கள்.

அமெரிக்காவில் பென்சில்வேனியா பல்கலைக்கழகத்தால் நடத்தப்படும் மொழியியல் தரவு கூட்டமைப்பு (LDC – Linguistic Data Consortium) ஒரு சிறந்த முன்மாதிரி எடுத்துக்காட்டு. இது ஆராய்ச்சி மற்றும் மேம்பாட்டு நோக்கங்களுக்காக பேச்சு மற்றும் உரை தரவுத்தளங்கள், சொற்களஞ்சியம் மற்றும் பிற ஆதாரங்களை உருவாக்குகிறது, சேகரிக்கிறது மற்றும் விநியோகிக்கிறது, இது பல்கலைக்கழகங்கள், நிறுவனங்கள் மற்றும் அரசு ஆய்வுக்கூடங்களின் திறந்த கூட்டமைப்பு ஆகும். ஆகவே இதேபோன்று இந்திய மொழிகளில் மொழியியல் தரவுக் கூட்டமைப்பு (LDCIL – Linguistic Data Consortium in Indian Languages) ஒன்றை அமைக்க திட்டமிடப்பட்டு வருவதாகச் சொல்கிறார்கள்.

சிறுபான்மை மொழி பொறியியல் செயல்படுத்தல் திட்டத்தின் (Enabling Minority Language Engineering) கீழ் எமிலி (EMILLE) மொழித்தொகுப்பு இங்கிலாந்தின் லங்காஸ்டர் பல்கலைக்கழகம் மற்றும் இந்திய மொழிகள் மத்தியக் கழகம் (CIIL) ஆகியவற்றுக்கிடையே ஒரு கூட்டு முயற்சியாக 2003 இல் உருவாக்கப்பட்டது.

எமிலி தமிழ் உரைத்தொகுப்பில் சுமார் 20 மில்லியன் சொற்கள் உள்ளன. தினகரன் இணைய தளத்திலிருந்து திரைப்படங்கள் பற்றிய செய்திகள் சுமார் ஒரு மில்லியன் சொற்கள். செய்திகள் சுமார் 9 மில்லியன் சொற்கள். மற்ற கட்டுரைகள் சுமார் ஒரு மில்லியன் சொற்கள். அரசியல் பற்றிய செய்திகளும் மற்றும் விமர்சனங்களும் சுமார் 5 மில்லியன் சொற்கள். விளையாட்டு பற்றிய செய்திகளும் மற்றும் விமர்சனங்களும் சுமார் ஒரு மில்லியன் சொற்கள். மைசூர் மத்திய இந்திய மொழிகள் கழகத்தின் மூலம் 3 மில்லியன் சொற்கள். இது ஐரோப்பிய மொழி வளங்கள் சங்கத்தால், இலாப நோக்கற்ற ஆய்வுகளில் பயன்படுத்த இலவசமாக விநியோகிக்கப்படுகிறது

ஐரோப்பா செக் குடியரசில் ப்ராக் நகரில் உள்ள சார்லஸ் பல்கலைக்கழகத்தில் லோகநாதன் ராமசாமி [வாக்கியங்களை நேரமைத்த ஒரு ஆங்கிலம் தமிழ் இருமொழித் தொகுப்பு தயாரித்தார்.](#)

என்டாம் (EnTam) என்ற பெயரிலுள்ள இந்த மொழித்தொகுப்பில் விவிலிய நூல், திரைப்படங்கள் மற்றும் செய்திகள் ஆகியவற்றிலிருந்து சுமார் 170,000 வாக்கியங்கள் உள்ளன.

தமிழ் விக்கிப்பீடியா உரைத்தொகுப்பு

விக்கிப்பீடியா மற்றும் அதன் குழு தளங்களில் இருந்து நான் [ஒரு தமிழ் உரைத்தொகுப்பு தயார் செய்துள்ளேன்](#). இந்த உரைத்தொகுப்பில் சுமார் 6 மில்லியன் சொற்கள் உள்ளன. தமிழ் விக்கிப்பீடியாவிலிருந்து சுமார் 3.8 மில்லியன் சொற்கள். இவை அறிவியல், கணிதம், சமூகம், தமிழ், தொழினுட்பம், நபர்கள், பண்பாடு, புவிவியல், வரலாறு ஆகிய பகுப்புகளில் உள்ளன. விக்கிமூலத்திலிருந்து, தற்காலத் தமிழ் எழுத்தாளர் உரைநடைப் படைப்புகள் மட்டும், சுமார் 1.7 மில்லியன் சொற்கள். இவை அண்ணாதுரை – ரங்கோன் ராதா, மு. வரதராசன் – அகல் விளக்கு, ரெ. கார்த்திகேசு – அந்திம காலம், கல்கி – பொன்னியின் செல்வன், கல்கி – பார்த்திபன் கனவு, கல்கி – சிவகாமியின் சபதம், கல்கி – அலை ஓசை, கல்கி – தியாக பூமி, கல்கி – மற்ற நாவல்கள், கல்கி – சிறுகதைகள், புதுமைப்பித்தன் – சிறுகதைகள், ஜெயகாந்தன் – சிறுகதைகள், பல ஆசிரியர்கள் – சிறுகதைகள், பல ஆசிரியர்கள் – கட்டுரைகள் ஆகிய பகுப்புகளில் உள்ளன. விக்கி செய்தியிலிருந்து சுமார் 0.4 மில்லியன் சொற்கள் மற்றும் விக்கி நூல்களிலிருந்து சுமார் 17 ஆயிரம் சொற்கள் உள்ளன. எவரும் இதை எளிதாகப் பதிவிறக்கம் செய்து பயன்படுத்த முடியும்.

பதிப்பாசிரியர்களால் தொகுக்கப்பட்ட உரைகளே மொழித்தொகுப்புக்குச் சிறந்தவை. இவை நல்ல வாக்கிய அமைப்பு கொண்டு இலக்கணப்படியும், பிழையற்றதாகவும் இருக்கும். மேலும் இவை நடப்பு மொழியில் இருக்க வேண்டும். நடைமுறையில் இல்லாத பழங்கால உரைகளாகவோ அல்லது கவிதைகளாகவோ இருக்கக்கூடாது. ஏனெனில் கவிதைகள் உரை விதிமுறைப்படி அமைவதில்லை. நமக்கு முழு வாக்கியங்கள்தான் தேவை. தலைப்புகள், பட்டியல்கள், அட்டவணைகள் ஆகியவற்றில் வாக்கியத் துண்டுகள் இருந்தால் பார்த்து நீக்கிவிடவேண்டும். வேற்று மொழிச்சொற்கள், கணித சூத்திரங்கள் போன்றவை வரும் வாக்கியங்களையும் நீக்கிவிடவேண்டும்.

உரைத்தொகுப்பு பகுப்பாய்வு செய்யும் கருவி

உரைத்தொகுப்புகள் மில்லியன் கணக்கான சொற்கள் அடங்கியவை. அவற்றைப் படிக்கவும், அவற்றில் சொற்களைத் தேடவும் மற்றும் ஆய்வு செய்யவும் சிறப்பு மென்பொருள் கருவிகள் தேவை. ஆங்கில திறந்த மூல உரைத்தொகுப்புக் கருவிகள் எதுவும் தமிழைச் சரியாகக் கையாளவில்லை. ஆனால் [பிரெஞ்சு குழு உருவாக்கிய TXM](#) தமிழுக்கு நன்றாக வேலை செய்கிறது. இது திறந்த மூல மென்பொருள். லினக்ஸ், மேக் மற்றும் விண்டோஸ் கணினிகளில் வேலை செய்யும். மிகப்பெரிய உரைத்தொகுப்புகளையும் கையாள இயலும். மொத்த சொற்குறிகள் (சொற்களும் நிறுத்தற்குறிகளும்) எண்ணிக்கை, மொத்த சொற்கள் எண்ணிக்கை, தனித்தன்மை வாய்ந்த சொற்குறிகள் எண்ணிக்கை, தொகுப்பில் ஒவ்வொரு சொல்லின் எண்ணிக்கை முதலிய புள்ளிவிவரங்களைத் தரும். மேலும் சொற்களும் அவற்றின் சூழலும் பற்றிய பகுப்பாய்வு (Concordance Analysis), இரண்டு சொற்கள் இணை நிகழ்வு (Cooccurrences) போன்ற பல வேலைகளையும் செய்ய முடியும்.

18. சொல்வகைக் குறியீடும் குறியிட்ட உரைத்தொகுப்புகளும்

பேச்சறிதல், இயற்கை மொழி பாகுபடுத்தல், தகவல் பெறுதல் மற்றும் தகவல் பிரித்தெடுத்தல் போன்ற இயல்மொழி செயலிகளில் குறியீடுகள் முக்கிய பங்கை வகிக்கின்றன.

ஆங்கிலத்தில் பொதுவாக ஒன்பது சொல்வகைகள் உள்ளன என்று பள்ளியில் கற்பிக்கின்றனர்: பெயர்ச்சொல் (noun), வினைச்சொல் (verb), சுட்டிடைச் சொல் (article), பெயருரிச்சொல் அல்லது பெயரடை (adjective), முன்னிடைச்சொல் (preposition), பதிலிடு பெயர் (pronoun), வினையுரிச்சொல் அல்லது வினையடை (adverb), இணையிடைச்சொல் (conjunction), மற்றும் வியப்பிடைச் சொல் (interjection). ஆங்கில சுட்டிடைச் சொல்லில் இரு உட்பிரிவினைகள் உண்டு. 'a', 'an' இரண்டும் வரையறாச் சுட்டிடைச் சொற்கள் (indefinite article). தமிழில் இவற்றுக்கு ஒப்பானவை 'ஒரு', 'ஓர்' எனலாம். 'the' வரையறு சுட்டிடைச் சொல் (definite article). தமிழில் வரையறு சுட்டிடைச் சொல் கிடையாது. ஆங்கிலத்தில் 'with', 'at', 'from', 'into' போன்றவை முன்னிடைச்சொற்கள். தமிழில் பின்னிடைச்சொற்கள்தான், 'போது', 'கூடிய', 'முதல்', 'வரை' போன்றவை.

சொல்வகைக் குறியீடு தொகுப்புகள் (Tagset)

தமிழ் இலக்கண நோக்கில் சொற்கள் பொதுவாக நான்கு வகைப்படும். அவை பெயர்ச்சொல், வினைச்சொல், இடைச்சொல், உரிச்சொல் ஆகும்.

மொழி வளரும்போது புதிய சொற்கள் சேர்ந்து கொண்டதானிருக்கும். பத்தாண்டுகளுக்கு முன் 'திறன்பேசி' என்ற சொல் கிடையாது. இம்மாதிரி பெயர்ச்சொல்தான் நிறைய சேரும். வினைச்சொற்கள் கொஞ்சம் சேரலாம். இவற்றைத் திறந்த சொல்வகை என்று சொல்கிறோம். "ஆனால்", "அல்லது", "மற்றும்" போன்ற சொற்கள் மேலும் சேர வழியில்லை. இவற்றை மூடிய சொல்வகை என்று சொல்கிறோம்.

தமிழ் மொழிக்கு இந்தியத் தரக் கட்டுப்பாட்டுச் செயலகம் (Bureau of Indian Standards – BIS) ஒரு குறியீடு தொகுப்பு (Tagset) வெளியிட்டுள்ளது. [இந்தியத் தரக் கட்டுப்பாட்டுச் செயலகம் வெளியிட்ட தமிழ் மொழிக்கான படிநிலை முறை குறியீடு தொகுப்பு இங்கே](#). இதில் மேல் மட்டத்தில் 11 வகைகள் உள்ளன. ஒவ்வொரு மேல் மட்ட வகைக்கும் பல உள்வகைகள் உள்ளன. எடுத்துக்காட்டாக, பெயர்ச்சொல்லில் 3 உள்வகைகள் உள்ளன – பொதுப்பெயர், தனிப்பெயர், கால இடப்பெயர். வினைச்சொல்லிலோ 3 மட்டங்களில் உள்வகைகள் உள்ளன. எடுத்துக்காட்டாக, முதல் மட்டத்தில் முதன்மை வினைச்சொல். அடுத்த படியில் வினை அடி, வினை முற்று, வினை எச்சம். அதற்கும் அடுத்த படியில் பெயர் எச்சம், நிபந்தனை எச்சம். ஆக மொத்தம் 45 குறியீடுகள் உள்ளன.

[அமிர்தா குறிச்சொல் தொகுப்பை இங்கே காணலாம்](#). இதில் 30 சொல்வகைகள் உள்ளன. எனினும் உங்கள் தேவையைப் பொருத்து சில எளிய குறியீடு தொகுப்புகளையும் பயன்படுத்தலாம். [பொதுக் குறியீடு தொகுப்பு \(Universal POS tags\)](#) என்பது அனைத்து மொழிகளிலும் ஒரு பெரிய அளவிலான பொதுக் குறியீடுகளைக் கொண்டது. நடைமுறையில் மிகவும் பயனுள்ளதாக இருக்கும் சொல் வகைக் குறியீடுகளை மட்டுமே இதில் சேர்த்துள்ளார்கள். இதில் 17 சொல் வகைகள் மட்டுமே உள்ளன. திறந்த சொல்வகைகள் 6, மூடிய சொல்வகைகள் 8, மற்றவை 3.

தமிழ், மற்ற பல இந்திய மொழிகளைப் போலவே, உருபனியல் அல்லது சொற்பகுப்பியல்படி வளமான (morphologically rich) மொழி ஆகும். பெரும்பாலும் நாம் ஒரு சொல்லை பெயர்ச்சொல்லா, பெயருரிச்சொல்லா, வினையுரிச்சொல்லா அல்லது பின்னிடைச்சொல்லா என்பதை சொற்றொடரியல் செயல்பாடு அல்லது சூழலை வைத்துத்தான் முடிவு செய்ய இயலும். ஆகையால் தமிழில் சொல் வகைக் குறியீடு செய்வது கொஞ்சம் சிக்கலான வேலையாக உள்ளது.

எடுத்துக்காட்டாக, “சென்னையில் நான்கு ஆண்டுகளாக வசிக்கிறேன்” இந்த வாக்கியத்தில், ‘ஆண்டுகளாக’ என்ற சொல்லை ‘ஆண்டு’ (பெயர்ச்சொல்) + ‘ஆக’ (வினையடை பின்னொட்டு) எனப் பிரிக்கலாம். தமிழ் மொழியில் ‘ஆக’ பொதுவாக ஒரு பெயர்ச்சொல்லுடன் இணைந்து ஒரு வினையடை உருவாக்குகிறது. எடுத்துக்காட்டு, “இந்தக் கட்டடம் பள்ளிக்கூடமாக இருக்கிறது”. ஆனால் இந்த வழக்கில் ‘ஆக’ என்பது ஆங்கிலத்தில் ‘for’ என்ற பொருளை ஒத்துள்ளது.

LEXICAL AMBIGUITY IN TAMIL.

□ Assign POS tags to words in a sentence considering its lexical ambiguity.

கோயிலில் ஆறு அடி உயரமான மணி உள்ளது .

NN	NN	NN	ADJ	NN	VF
NN	CRD	VF	ADJ	NNP	VF

<Six feet tall bell is in the temple>

CEN Amrita Vishwa Vidyapeetham Coimbatore 10

சொல் தெளிவின்மை எடுத்துக்காட்டு

ஆங்கில சொல்வகைக் குறியிடும் கருவிகள் (POS taggers)

குறியீடு செய்யும் கருவிகள் இரண்டு வகையானவை. ஆங்கிலத்தில் முதலில் விதிமுறைகள்படி குறியீடு செய்யும் கருவிகள்தான் வந்தன. பின்னர்தான் புள்ளி விவரங்கள்படியான கருவிகள் வந்தன. புள்ளிவிவர அடிப்படையிலானவை அண்டை சொற்களைப் பற்றிய தகவலைப் பயன்படுத்துகின்றன.

தமிழ் சொல்வகைக் குறியிடும் கருவிகள்

சொல்வகைக் குறியீடு (POS tagging) செய்வதன் ஒரு நோக்கம் பல்பொருள் ஒரு சொல் (homonym) போன்ற தெளிவிலா உரையைத் தெளிவுபடுத்துவதாகும். குறியீடு செய்யும் கருவிகளை விதிமுறை அடிப்படை, புள்ளிவிவர அடிப்படை என்று இரண்டாக வகைப்படுத்தலாம். விதிமுறை அடிப்படையிலானவை குறியீடு தெளிவின்மையை வேறுபடுத்த கைமுறையாக எழுதப்பட்ட விதிகளைப் பயன்படுத்துகின்றன.

வழக்கமாக, புள்ளிவிவர மாதிரிகள் அண்டை சொற்களைப் பற்றிய தகவலைப் பயன்படுத்திப் பொருத்தமான குறியீடுகளை இடுகின்றன. எடுத்துக்காட்டாக, ‘நான் படி ஏறினேன்’ என்ற வாக்கியத்தில் ‘படி’ என்பது படிக்கட்டு என்றால் பெயர்ச்சொல், அல்லது படிக்கச் சொன்னால் வினைச்சொல். இந்த வாக்கியத்தில் ஒரு வினைச்சொல்லுக்கு முன்னால் இருப்பதால் அது ஒரு பெயர்ச்சொல்லாகத்தான் இருக்க வேண்டும்.

தமிழ் ஒட்டு நிலை மொழி என்பதால் உருபனியல் ஆய்வு செய்தே குறியீடு செய்யலாம். வாசு ரெங்கநாதன் தயாரித்த, ஓரளவு சொற்களை மட்டுமே குறியீடு செய்யும், உருபனியல் குறியீட்டுக் கருவி ஒன்று உள்ளது. தமிழ்மொழி போன்ற ஒட்டு நிலை மற்றும் கட்டற்ற சொல் வரிசை மொழிகளுக்கான சிறந்த குறியீடுதல் திட்டத்தை அடையாளம் காண, குறியீடுதல் வழிமுறைகளைப் பல்வேறு கோணங்களிலிருந்து முயற்சிக்க வேண்டும்.

2016 ஆம் ஆண்டு மே மாதம் அண்ணா பல்கலைக்கழக KBC ஆய்வுமையம் தமிழ் சொல்வகைக் குறியிடும் கருவி ஒன்றை வெளியிட்டது. இது GNU GPL 3.0 பொது உரிமத்தின் கீழ் வெளியிடப்பட்டது. பதிவிறக்கம் கிடையாது, எழுதிக் கேட்க வேண்டும்.

RDRPOSTagger என்ற கட்டற்ற சொல்வகைக் குறியீடு செய்யும் மென்பொருளைப் (POS tagger) பயன்படுத்தி நான் ஒரு தமிழ் சொல்வகைக் குறியிடும் கருவி தயாரித்தேன். இந்தக் கருவி [30 சொல்வகைகள் கொண்ட அமிர்தா குறிச்சொல் தொகுப்பு \(Amrita Tagset\)](#) படி குறியீடு செய்யத் தயார்

செய்தது. கைமுறையாகக் குறியீடு செய்த சொல்வங்கி (tagged corpus) பயன்படுத்தி, இயந்திரக் கற்றல் (machine learning) மூலம் இது தயார் செய்தது. நீங்கள் இதை வைத்து உங்கள் தமிழ் உரையைக் கட்டளை வரியிலும் (command line) பைதான் நிரல் மூலமும் குறியீடு செய்யலாம். மாதிரி குறியீடு செய்த வாக்கியம் இங்கே:

என்ன/QW செய்வது/VBG என்று/COM அவர்களுக்குத்/PRP தெரியவில்லை/VAX ./.

கைமுறையாக சொல்வகைக் குறியீடு செய்த உரைத்தொகுப்புகள்

2016 ஆம் ஆண்டு மே மாதம் அண்ணா பல்கலைக்கழக KBC ஆய்வுமையம் சொல்வகைக் குறியீடு செய்த உரைத்தொகுப்பு ஒன்றை வெளியிட்டது. இது அரை மில்லியன் சொற்கள் கொண்ட இந்திய மொழிகளில் கிடைக்கக்கூடிய மிகப்பெரிய கைமுறையான சொல்வகைக் குறியீடு செய்த உரைத்தொகுப்பு ஆகும். இது கல்கி கிருஷ்ணமூர்த்தியால் எழுதப்பட்ட புகழ்பெற்ற 20 ஆம் நூற்றாண்டு தமிழ் நாவல் “பொன்னியின் செல்வன்”. இதில் சுமார் 50,000 வாக்கியங்கள் உள்ளன. இந்தியத் தரக் கட்டுப்பாட்டுச் செயலகம் வெளியிட்டு மற்றும் தமிழ் மெய்நிகர் கல்விக்கழகமும் அங்கீகாரம் செய்த குறியீடு தொகுப்பை வைத்து இந்தக் குறியீடு செய்யப்பட்டது.

19. வாக்கியக் கூறு பிரித்தலும், பெயரிட்ட உருபொருள் அடையாளம் காணுதலும்

நாம் எண்ணங்களை சொற்களாலும் வாக்கியங்களாலும் வெளிப்படுத்துகிறோம். எல்லா மொழிகளும் சொற்களையும் வாக்கியங்களையும் கொண்டிருக்கின்றன. ஆனால் அவை பல்வேறு வழிகளில் ஒழுங்கமைக்கப்படுகின்றன.

தொடரியல் (syntax)

தொடரியல் என்பது சொற்களை வைத்து எவ்வாறு வாக்கியங்களை அமைக்கிறோம் என்ற வாக்கியக் கட்டமைப்பு ஆய்வு. தமிழ் இலக்கணப்படி எழுவாய் என்பது ஒரு வாக்கியத்தில் செயலைக் காட்டும் சொல்மீது “யார், எது, எவை” என வினவும் போது கிடைக்கும் பதில் ஆகும். செயப்படுபொருள் என்பது “யாரை, எதை, எவற்றை” என்பதின் பதில் ஆகும். பொருள் முடிந்து நிற்கும் வினைச் சொல் நிலை பயனிலை எனப்படுகிறது. ஆங்கிலத்தில் எளிய வாக்கியங்களில் பொதுவாக எழுவாய் (subject) முதலில் வரும், அடுத்து பயனிலை என்னும் வினைச்சொல் (verb), கடைசியாக செயப்படுபொருள் (object). *I went home* (S-V-O order). மாறாக தமிழில் எளிய வாக்கியங்களில் பொதுவாக எழுவாய் முதலில் வரும், அடுத்து செயப்படுபொருள், கடைசியாக பயனிலை என்னும் வினைச்சொல். நான் வீட்டுக்குப் போனேன் (S-O-V order). இதைத் தமிழில் எ-செ-ப ஒழுங்கு என்று சொல்கிறோம்.

வாக்கியக் கூறாக்கம் (chunking or shallow parsing)

ஒரு வாக்கியத்தில் இரண்டு சொற்களுக்கு இடையில் உள்ள உறவை நீங்கள் கண்டுபிடிக்க விரும்புகிறீர்கள் என்று வைத்துக்கொள்வோம். இந்த வாக்கியத்தைப் பல கிளைகளுள்ள மரம் போல் உருவாக்கலாம். பாகுபடுத்தி என்னும் கருவியைப் பயன்படுத்தி இந்தக் கிளைப் படத்தை உருவாக்கலாம். ஸ்டான்ஃபோர்ட் பாகுபடுத்தி (Stanford parser) கருவியைப் பயன்படுத்தி ஆங்கில வாக்கியத்தில் சொற்களுக்கு இடையே இலக்கண உறவுகளின் வரைகலை உருவகிப்பை உருவாக்க முடியும்.

இயல் மொழி ஆய்வில் சொல் வகைக் குறியீடு செய்வதால் மட்டுமே உரைகளின் உண்மையான பொருளை சில நேரங்களில் அறிய முடியாது. அதே நேரத்தில் மேற்கண்டபடி பெருமுயற்சி செய்து வாக்கியத்தின் எல்லாக் கிளைகளையும் உருவமைப்புப்படி பிரித்தெடுக்கவும் தேவையில்லை. இவற்றுக்கு இடையில் வாக்கியக் கூறாக்கம் செய்தால் போதும். சொல்வகைக் குறியீடு செய்தபின் நமக்கு “வள்ளுவர்” மற்றும் “கோட்டம்” இரண்டும் பெயர்ச்சொற்கள் என்று தெரியவரும். தனிச் சொற்களுக்குப் பதிலாக “வள்ளுவர் கோட்டம்” என்ற சொற்றொடரை எடுத்தால்தான் எதைப் பற்றிய உரை என்று கூற இயலும் அல்லவா?

மற்ற மொழிகளுடன் ஒப்பிடும்போது தமிழில் சொல் வரிசைகண்டிப்பான விதிமுறைகள்படி அமைவதில்லை. சொற்கள் முன்னுக்குப் பின் ஓரளவு மாறி வரலாம் (relatively free word order language). ஆனால் வாக்கியக் கூறுகள் (phrases and clauses) அளவில் பார்க்கும்போது சொல் வரிசை கண்டிப்பான விதிமுறைப்படிதான் அமைகிறது.

வாக்கியங்களிலிருந்து சொற்றொடர்களைப் பிரித்தெடுப்பதை வாக்கியக் கூறாக்கம் (sentence chunking) என்று சொல்கிறோம். இதைச் செய்யும் கருவிக்கு மேலோட்டமான பாகுபடுத்தி (shallow parser) என்று பெயர். இக்கருவி சொல் வகைக் குறியீடு செய்த உரையை எடுத்துக் கூறாக்கம் செய்கிறது. சொல் வகைக் குறியீடு போலவே வாக்கியக் கூறுகளிலும் பெயர்ச்சொற்றொடர், வினைச்சொற்றொடர், முன்னிடைச்சொற்றொடர் போன்ற குறியீடுகள் உள்ளன. பெயரிட்ட உருபொருள் அடையாளம் (Named Entity Recognition) காண இந்த வாக்கியக் கூறாக்கம் மிகவும் உதவியானது.

பெயர்ச்சொற்றொடர் கூறாக்கம் உருவாக்க, சொல்வகைக் குறியீடுகளைப் பயன்படுத்துகிறோம். ஆங்கிலத்தில் ஒரு எளிய எடுத்துக்காட்டைப் பார்ப்போம். ஒருக்கால் a, an, the போன்ற சுட்டிடைச்

சொல்லில் தொடங்கி, எத்தனை பெயரடைகள் வேண்டுமானாலும் இருந்து, பின்னர் பெயர்ச்சொல் வந்தால் பெயர்ச்சொற்றொடரை (NP) உருவாக்க வேண்டும் என்று விதி கூறுகிறது. இதைத் தேட ஒரு தேடுகுறித்தொடர் (regular expression or Regex) விதிமுறையை உருவாக்க முடியும்.

சொல்வகைக் குறியீடு செய்த உரையை உள்ளீடாகக் கொடுத்து அடுத்த செயல்முறையாக வாக்கியங்களைக் கூறுகளாகப் பிரித்து வாங்குகிறோம். சொல்வகைக் குறியீடு போலவே இதற்கும் வினைச்சொல் சொற்றொடர் (verb phrase – VP) பெயர்ச்சொல் சொற்றொடர் (noun phrase – NP) போன்ற தரமான குறியீடு தொகுப்புகள் உள்ளன.

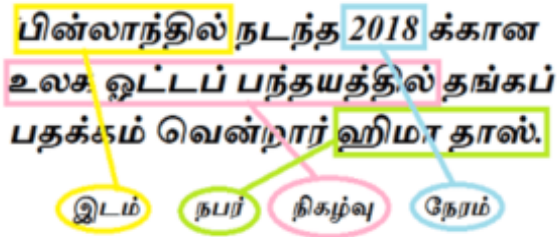
“the little yellow dog barked at the cat” என்ற வாக்கியத்தை எடுத்துக்கொள்வோம்.

இந்த விதியை வைத்துக் கீழ்க்கண்ட இரண்டு பெயர்ச்சொற்றொடர்களைப் (noun phrase NP) பிரித்தெடுக்கலாம்

(NP the/DT little/JJ yellow/JJ dog/NN)

(NP the/DT cat/NN).

[ஐதராபாதி](#)லுள்ள இந்திய தகவல் தொழில்நுட்பக் கழகத்தின் இணையதளத்தில் தமிழுக்கும் மற்ற பல இந்திய மொழிகளுக்கும் ஒரு மேலோட்டமான பாகுபடுத்தியை (shallow parser) தரவிறக்கம் செய்யலாம். அனைத்து வெளியீடுகளும் சக்தி தரநிலை வடிவமைப்பில் (Shakti Standard Format – SSF) உள்ளன. இதை அவர்களுடைய இணையதளத்திலேயே ஓட்டியும் பார்க்கலாம்.



பெயரிட்ட உருபொருள் அடையாளம் காணுதல்

பெயரிட்ட உருபொருள் அடையாளம் காணுதல்

கடந்த பத்தாண்டுகளாக வலைத்தளங்கள், வலைப்பதிவுகள், மின்னஞ்சல், அரட்டைகள் போன்ற பல்வேறு ஊடக வகைகள் குறிப்பிடத்தக்க அளவு அதிகரித்துள்ளன. சிறு நகரங்களில் இருந்து உருவாக்கப்படும் உள்ளடக்கங்கள் அதிகரித்து வருகின்றன. குறிப்பாக வணிக நிறுவனங்கள் தங்கள் தயாரிப்புகளிலும் சேவைகளிலும் இப்பெரிய தரவுகளைப் பயன்படுத்தி, பொதுமக்களுடைய பின்னூட்டங்களைத் தெரிந்துகொள்ள ஆர்வமாக உள்ளன.

தனிப்பெயர்ச்சொற்களை கடல், மலை, ஆறுகள், கண்டங்கள் போன்ற புவியியல் அமைப்புகள், நாடுகள், மாநிலங்கள் போன்ற அரசியல் அமைப்புகள், திங்கட்கிழமை போன்ற காலநிலை மற்றும் நிறுவனங்கள், நபர்கள் முதலியனவாக அடையாளம் கண்டு குறியிடுதலை பெயரிட்ட உருபொருள் அடையாளம் காணுதல் என்று கூறுகிறோம்.

இயல்மொழி ஆய்வு சொல்வழக்கில் பெரும்பாலும் “தனிப்பெயர்ச்சொற்கள்” (Proper Nouns) என்பதற்குப் பதிலாக “பெயரிட்ட உருபொருட்கள்” (Named Entities) என்ற கருத்தைப் பயன்படுத்துகிறோம். பெரும்பாலும் இது ஒன்றுக்கு மேற்பட்ட சொற்களைக் கொண்டுள்ளது.

ஆங்கிலத்தில் பெயரிட்ட உருபொருள் அடையாளம் காணும் கருவிகளைப் பழக்குவதற்காகவே ஜிஎம்பி மொழித் தொகுப்பிலிருந்து ஒரு பகுதியை எடுத்துக் குறியீடு செய்து தயாரித்துள்ளார்கள். ஸ்டான்போர்ட் பெயரிட்ட உருபொருள் (NER) குறியீடு செய்யும் கருவி ஆங்கில மொழிக்கு மட்டுமே வடிவமைக்கப்பட்டுள்ளது. [NLTK \(Natural Language Toolkit\)](#) மற்றும் [ஸ்டான்போர்ட்](#)

NER குறியீடு செய்யும் கருவியில் இயந்திரக் கற்றல் பயிற்சி அளித்து மற்றொரு மொழிக்கு மாதிரியைத் தயாரித்தது மட்டுமல்லாமல் அதற்கான படிகளையும் விவரமாக இங்கே ஒருவர் எழுதியிருக்கிறார். நல்ல தரமான பெயரிட்ட உருபொருள் குறியிட்ட தமிழ் உரை கிடைத்தால் இதே படிக்களைப் பின்பற்றி தமிழுக்கும் பயிற்சி அளிக்க முடியுமா என்று முயற்சி செய்து பார்க்கலாம்.

நடைமுறை எடுத்துக்காட்டு: வாடிக்கையாளர்கள் புகார்களை வகைப்படுத்தி உடன் நடவடிக்கை எடுத்தல்

கட்டமைக்கப்படாத இயல்மொழி உரைகளில் மதிப்பு மிக்க தகவல்கள் நிறைந்துள்ளன. ஆனால் பொருத்தமானது என்னவென்பதைக் கண்டுபிடிப்பது எப்போதும் ஒரு சவாலான பணியாகும். சமூக ஊடகங்கள், மின்னஞ்சல், வலைப்பதிவுகள், செய்திகள் மற்றும் கல்விக் கட்டுரைகளில் இருந்து வரும் விரிவான தரவுகளைப் பாருங்கள். அவற்றிலிருந்து மதிப்பு மிக்க தகவல்களைப் பிரித்தெடுத்தப்படும், வகைப்படுத்துவதும் மற்றும் கற்றுக்கொள்வதும் மிகவும் கடினமானது ஆனால் மிகவும் அவசியமானது.

வாடிக்கையாளர்களின் கருத்தை சுமுகமான முறையில் கையாளுவதற்குப் பல வழிகள் உள்ளன. பெயரிட்ட உருபொருள் அடையாளம் காணுதல் அவற்றில் ஒன்று. இந்த செயல்முறையைப் புரிந்து கொள்வதற்கு இதோ ஒரு எடுத்துக்காட்டு. நீங்கள் உலகளாவிய பல கிளைகள் கொண்ட ஒரு கடையின் வாடிக்கையாளர் ஆதரவுத் துறைக்கு மேலாளராக இருக்கிறீர்கள் என்று வைத்துக் கொள்வோம். நூற்றுக்கணக்கான மின்னஞ்சல்கள் வருகின்றன. உங்கள் வாடிக்கையாளர்களின் கருத்துகளில் பல குறிப்புகளை நீங்கள் காணலாம். எடுத்துக்காட்டாக, வாடிக்கையாளர் மின்னஞ்சலிலிருந்து, நீங்கள் பெயரிட்ட உருபொருள் அடையாளம் கண்டால், அது மயிலாடுதுறை என்ற இடம் பற்றியது மற்றும் தேங்காய் எண்ணெய் என்ற தயாரிப்பு பற்றியது என்று தெரியவரலாம். புகாரை வகைப்படுத்தவும் நிறுவனத்திற்குள்ளேயே அதைக் கையாளும் கிளைக்கோ அல்லது சம்பந்தப்பட்ட துறைக்கோ அதை ஒதுக்கவும் இந்தத் தகவலைப் பயன்படுத்தலாம்.

20. தமிழின் தனித்தன்மைகளை வைத்துக் குறியிட்ட உரைகள் தேவையைக் குறைக்க முடியுமா?

சொல்வகைக் குறியீடு ஒரு சவால் மிகுந்த சிக்கலான பணியாகும். ஏனெனில் அகராதியில் இல்லாத தனிப்பெயர்ச்சொற்கள், மற்ற மொழிச் சொற்கள், மாற்று எழுத்துக்கோர்வை, எழுத்துப் பிழைகள், தெரியாத சொற்கள் போன்றவை வரலாம்.

இயந்திரக் கற்றல் நுட்பங்களைப் பயன்படுத்தி ஆங்கிலத்துக்குப் பல சொல்வகைக் குறியீடு செய்யும் கருவிகள் உருவாக்கப்பட்டுள்ளன. இயந்திரக் கற்றல் நுட்பங்களுடன், விதிகள் சார்ந்த அணுகுமுறைகளைக் கலந்தும் சில கருவிகள் உள்ளன. எனினும், பெரும்பாலானவை உருபனியல் அல்லது சொற்பகுப்பியல் உத்திகளில் கவனம் செலுத்துவதில்லை. அதற்குப் பதிலாக அவர்கள் மிகுதியான குறியீடு செய்த உரைத் தரவுகளையே நம்பியிருக்கிறார்கள். இந்த சொல்வகைக் குறியீடு செய்யும் கருவிகளின் துல்லியம் 93 முதல் 98% வரை உள்ளது. ஆங்கிலத்துக்கு இது பொருத்தமான அணுகலாக இருக்கலாம். ஆனால் இந்திய மொழிகளில் வளங்கள் குறைவாக இருப்பதால் இது உகந்த வழியல்ல. மேலும் பொதுவாக இந்திய மொழிகளை உருவமைப்புப்படி வளமான மொழிகள் என்று சொல்லலாம். இது சில புதிய சிக்கல்களை உண்டாக்குகிறது. இது தவிர, இந்திய மொழிகளுக்கு வாக்கியத்தில் சொல்வரிசை ஒப்பீட்டளவில் கறாரான விதிமுறைப்படி அமைவதில்லை.

கைமுறையாகக் குறியீடு செய்ய செலவு அதிகம். குறைந்த துல்லியம் உள்ள கருவியால் முதலில் குறியீடு செய்து அதைக் கைமுறையாகத் திருத்துதல் வேலையைக் குறைக்கும். இருப்பினும் மொழியியலில் பயிற்சி பெற்றவர்கள்தான் இந்த வேலையைச் செய்ய இயலும்.

மும்பை இந்திய தொழில்நுட்பக் கழகம் இந்தி மொழியில் ஆய்வு

மும்பையிலுள்ள இந்திய தொழில்நுட்பக் கழகத்தினர் இந்தி மொழியில் ஆய்வு செய்து ஒரு ஆய்வறிக்கையை வெளிட்டனர். இந்த வழிமுறை 15 ஆயிரம் சொற்கள் கொண்ட சிறிய உரைத்தொகுப்பைப் பயன்படுத்துகிறது. அதிக சொற்களுள்ள அகராதி, கிளைகளாகப் பிரிந்து முடிவுக்கு வரும் கற்றல் வினைச்சரம், முழுமையான உருபனியல் பகுப்பாய்வு ஆகியவற்றை அடிப்படையாகக் கொண்டது. சொல்வகைக் குறியீடு செய்வதில் தற்போதைய துல்லியம் 93.45% ஆகும். இதை மேலும் மேம்படுத்தக்கூடும்.

உருபனியல்படி வளமான மொழிகளில் ஒரு சொல் பல உருபங்களைக் கொண்டிருக்கலாம். மேலும் அவற்றில் உருபன் எல்லைகளைக் கண்டுபிடிப்பதும் கடினம், ஏனென்றால் அவை ஒன்றாக இணைந்திருக்கும். மேலும் அவை பொதுவாக கட்டற்ற சொல் வரிசையைக் கொண்டுள்ளன. ஆகவே நிலையான சூழல் அமைப்புகளை புள்ளிவிவர அணுகுமுறைகளில் கற்பிக்க இயலாது.

தொடரியல் மற்றும் சொற்பகுப்பியல் பற்றிய விரிவான மொழியியல் பகுப்பாய்வு, பின்னொட்டுகளைக் கையாளுதல், துல்லியமாக வினைக் குழுக்களை அடையாளம் காணுதல் மற்றும் ஐயம்கற்றல் விதிகளைக் கற்றல் ஆகியவைதான் இந்த அமைப்புமுறையின் முக்கிய அம்சங்கள்.

இந்தி மொழியில் (குறிப்பாக பெயர்ச்சொற்களிலும் வினைச்சொற்களிலும்) ஒரு சொல்லின் சொல்வகைக் குறியீடுக்கு பின்னொட்டுகள் வலுவான சுட்டிக்காட்டிகள் ஆகும். எடுத்துக்காட்டாக, இந்தியில் 'ஜாயேகா' (போவேன்) என்னும் சொல். இது தோன்றுகின்ற சூழலின் அடிப்படை, மற்றும் இது 'ஜா' (போ) என்ற அடிச்சொல்லுடன் 'யேகா' என்ற பின்னொட்டு இணைந்தது என்பதை வைத்து இது வினைச்சொல் என்று தீர்மானிக்க முடியும். ஒரு சொல் பெயர்ச்சொல்லா அல்லது வினைச்சொல்லா என்று முடிவு செய்ய நாம் பயன்படுத்தும் அடிப்படைத் தத்துவம் ஒரு சொல் பெயரடையா அல்லது வினையடையா என்று முடிவு செய்யப் பயன்படுத்துவதைவிட முற்றிலும் வேறுபட்டது. எடுத்துக்காட்டாக, 'பர்' என்ற சொல் இணையிடைச்சொல், பெயர்ச்சொல் அல்லது பின்னிடைச்சொல்லாக இருக்கலாம். ஆகவே இது இணையிடைச்சொல்-பெயர்ச்சொல்-பின்னிடைச்சொல் என்ற ஐயம் திட்டத்தில் சேர்கிறது.

இந்த அணுகுமுறையைத் தமிழ் போன்ற ஒட்டுமுறை மொழிகளுக்குப் பயன்படுத்தலாம். தமிழுக்கான பின்னொட்டு மாற்று விதிகள் (suffix replacement rules – SRRs), சொல்வங்கி, குழு அடையாளம் மற்றும் உருபன் பகுப்பாய்வு விதிகளைத் தயார் செய்தால் போதும்.

மும்பை இந்திய தொழில்நுட்பக் கழகம் மராத்தி மொழியிலும் இதே ஆய்வு

மேற்கண்ட ஆய்வைத் தொடர்ந்து மராத்தி மொழியிலும் ஆய்வு செய்து இந்திய தொழில்நுட்பக் கழகத்தினர் ஒரு ஆய்வறிக்கையை வெளியிடப்பட்டனர். இதன் தலைப்பு [“வினைச்சொற்களில்தான் நுட்பம் அடங்கியுள்ளது: உருவமைப்புப்படி வளமான மொழிகளில் வாக்கியங்களைக் கூறுகளாகப் பிரித்ததில் கிடைத்த அனுபவங்கள்”](#).

உருவமைப்புப்படி வளமான மொழிகளில் (morphologically rich languages) மொழி பற்றிய நுணுக்கம் பயன்படுத்தி பெரிய அளவில் குறியீடு செய்த உரை தேவைப்படுவதைக் குறைக்க முடியும் என்பது இவர்களின் கூற்று. மராத்தி மொழியில் இருபதாயிரம் சொற்கள் மட்டுமே கொண்ட ஒரு சிறிய உரைத்தொகுப்பை வைத்து 94% சரியாக சொல்வகைக் குறியீடும் 97% சரியான வாக்கியங்களைக் கூறுகளாகப் பிரித்தலும் (shallow parsing or chunking) செய்து இதை நிரூபித்துள்ளனர்.

மராத்தி மொழியில் ‘ஜோடணாரா’ (connecting) என்ற சொல்லில் ‘ணாரா’ என்ற பின்னொட்டு வேறெந்தச் சொல் வகையிலும் வராது. ஆகவே இந்தப் பின்னொட்டை மட்டும் பார்த்தே நாம் இதை வினைச்சொல் என்று குறியீடு செய்து விடலாம். வாக்கியங்களைக் கூறுகளாகப் பிரிக்கும்போது, ஒரு வினையாலணையும் பெயர் கொண்டிருக்கும் வாக்கியக் கூறை எப்போதும் VGNN என்று குறியீடு செய்கிறோம். இதேபோல், ஒரு வினையெச்சம் கொண்டிருக்கும் வாக்கியக் கூறை எப்போதும் VGINF என்று குறியீடு செய்கிறோம். ஆகவே, வாக்கியத்தில் வினையாலணையும் பெயர் மற்றும் வினையெச்சத்தை சரியாக சொல்வகைக் குறியீடு செய்வது முக்கியமாகிறது. மராத்தி மொழியில் “ண்யாச்சா” என்ற பின்னொட்டு வினையாலணையும் பெயருக்கு மட்டுமே வரும். இதேபோல், பின்னொட்டு “யலா” வந்தால் எப்போதுமே அது வினையெச்சம்தான். இம்மாதிரி பின்னொட்டுத் தகவலைப் பயன்படுத்துவதால் ஒரு புள்ளிவிவர சொல்வகைப் படுத்தி முன்பு பார்த்தேயிராத சொற்களையும் குறியீடு செய்ய இயலும்.

ஒப்பீடு செய்ய சில எடுத்துக்காட்டுகளைப் பார்ப்போம். நான் திறந்த மூல RDR சொல்வகைக் குறியீடும் கருவியை தமிழுக்குப் பழக்கியபோது சுமார் 200 ஆயிரம் சொற்கள் கொண்ட குறியீடு செய்த உரைத்தொகுப்பைப் பயன்படுத்தினேன். 90% க்கும் கீழேதான் சரியான குறியீடுகள் செய்தது. வியட்நாம் மொழியில் 630 ஆயிரம் சொற்கள் பயன்படுத்தி இதே RDR சொல்வகைக் குறியீடும் கருவி 93.4% சரியான குறியீடுகள் செய்தது என்று சொன்னார்கள்.

ஆங்கிலம் அளவுக்கு இம்மொழிகளில் வளங்கள் கிடையாது. எனினும் மொழியின் செழுமையைப் பயன்படுத்தி இந்த வளப் பற்றாக்குறையை ஓரளவு ஈடுசெய்ய முடியும். குறிப்பாக, இது போன்ற மொழிகளில் சொல்வகை பற்றிய தகவல்களை பின்னொட்டுகள் நிறைய அடக்கியுள்ளன. இந்த நுட்பத்தை நாம் மேலோட்டமான பாகுபடுத்திக்கு (shallow parser) பயன்படுத்திக் கொள்ளலாம். மேலும், வினைச்சொற்களின் கட்டமைப்பு கண்டிப்பான விதிமுறைப்படிதான் அமைகிறது. முக்கிய வினைச்சொற்கள் மற்றும் துணை வினைச்சொற்களுக்கு இடையே உள்ள தெளிவின்மையைக் குறைக்க நாம் இதைப் பயன்படுத்த முடியும்.

எடுத்துக்காட்டாக, “உடண்யாச்சா” என்ற ஒரு சொல் வந்தால், பயிற்சி தரவில் இல்லாத சொல்லாக இருந்தாலும், பின்னொட்டுத் தகவலையும் பயன்படுத்தும் ஒரு வகைப்படுத்தி, அதன் அடிப்படையிலேயே அதை சரியாக அடையாளம் காண முடியும். ஆனால் புள்ளிவிவரங்களை மட்டுமே பயன்படுத்தும் ஒரு வகைப்படுத்தி அதை சரியாக அடையாளம் காட்டாது.

ஒரு சக்திவாய்ந்த உருபனியல் பகுப்பாய்வியும் (morphological analyzer) அதற்கு ஆதரவாக ஒரு அதிகமான சொற்கள் கொண்ட சொல்வங்கியும் (lexicon) மற்றும் ஒரு எளிய ஆனால் துல்லியமான வினைக் குழு அடையாளங்காட்டியும் (Verb Group Identifier) தேவை. இவற்றை வைத்து ஒரு துல்லியமான அதி நவீன வரிசை வகைப்படுத்தியை (sequence classifier) உருவாக்க முடியும்.

இதைத் தமிழுக்கு செயல்படுத்துவது எப்படி?

இம்மாதிரி மொழிகளுக்கு அவற்றின் அம்சங்களைப் பயன்படுத்தித்தான் ஒரு மிகவும் துல்லியமான வகைப்படுத்தியை உருவாக்க முடியும் என்று காட்டுகிறோம். ஆங்கிலத்தைப் பார்த்து அதே போல பெரிய அளவில் குறியீடு செய்த உரைகளை மட்டுமே நம்பினால் வேலைக்கு ஆகாது. ஆகவே இம்மாதிரி மொழிகளுக்கு உரைகளைக் கைமுறையாகக் குறியீடு செய்வதைவிட நல்ல உருபனியல் பகுப்பாய்வியில் முதலீடு செய்வதே உசிதமானது.

மொழியியலாளர்களுக்கும் நிரலாளர்களுக்கும் நெருங்கிய ஒத்துழைப்பு தேவை என்பதற்கு நல்ல எடுத்துக்காட்டு

மொழியியலாளர்களுக்கும் மென்பொருள் உருவாக்குநர்களுக்கும் இடையே நெருங்கிய ஒத்துழைப்புத் தேவை என்று முன்னொரு கட்டுரையில் கூறினோம். அது சரியென நிறுவ இது ஒரு நல்ல எடுத்துக்காட்டு. ஏனெனில், இதை செயற்படுத்த இரண்டு துறைகளிலுமே அறிவும் அனுபவமும் தேவை. எம்மாதிரி பின்னொட்டுகள் வந்தால் எந்த சொல்வகைக் குறியீடு போடலாம், மற்றும் அதற்கு விதிவிலக்குகள் யாவை என்று கணிக்க மொழியியல் திறன் அவசியம். இந்த வினைச்சரங்களையெல்லாம் நிரலில் எழுதி நன்றாக ஓட வைக்கத் திறமையான நிரலாளர்கள் தேவை.

21. சொல்வலையும் சொல்லின் பொருளில் ஐயமகற்றலும்

அடுத்து வரும் நான்கு வாக்கியங்களைப் பாருங்கள்.

- அவன் வீட்டிற்குச் சென்று இட்லி சாப்பிட்டான்.
- பின்னர் அவன் சமையலறையைச் சுத்தம் செய்து விட்டு நாற்காலியில் அமர்ந்தான்.
- சிறிது நேரம் கழித்து, அவன் இருக்கையில் இருந்து எழுந்தான்.
- அவன் படுக்கைக்குச் சென்றான், சில நிமிடங்களில் அவன் சத்தமாகக் குறட்டை விட்டான்.

இயல்மொழி செயலாக்கத்தில், வாக்கியங்களின் பொருளை அறிய கணினி நிரல்களைப் பயன்படுத்துகிறோம். மேலே உள்ள நான்கு வாக்கியங்களில், சொல்வலை (wordnet) உதவியுடன், கணினி நிரலால் பின்வருவனவற்றை அடையாளம் காண முடியும்.

- “இட்லி” என்பது ஒரு உணவு வகை.
- “சமையலறை” என்பது “வீடு” என்பதன் ஒரு பகுதி.
- “நாற்காலி” என்பது “இருக்கை” என்பதே.
- “குறட்டை” என்பது “தூக்கம்” என்பதைக் குறிப்பால் உணர்த்துகிறது.

சுருக்கமாக, சொல்வலை என்பது சொற்பொருள் உறவுகளால் ஒன்றாக இணைக்கப்பட்ட தமிழ்ச் சொற்களின் தரவுத்தளமாகும். இது மிகையேற்றிய அகராதி மற்றும் நிகண்டு போன்றது. ஒரு சொல்லின் பொருளையே தரும் இன்னொரு சொல்லை நேர்ச்சொல் (synonym) என்கிறோம். இம்மாதிரி ஒரு பொருளைத் தரும் பல சொற்களைத் திரட்டினால் அதை நேர்ச்சொல் தொகுப்பு (synonym set or synset) எனலாம். பெயர்ச்சொற்கள், வினைச்சொற்கள், பெயரடை மற்றும் வினையடை சொற்கள் நேர்ச்சொல் தொகுப்புகளாகப் (synsets) பிரிக்கப்படுகின்றன. ஒவ்வொரு நேர்ச்சொல் தொகுப்பும் ஒரு தனித்துவமான கருத்தை வெளிப்படுத்துகிறது.

பழந்தமிழரின் சொற்களின் தரவுத்தளங்கள்

உரிச்சொல் அல்லது உரிச்சொற் பனுவல் என்பவை சொற்களுக்கு விளக்கம் தரும் நூல்களைக் குறிக்கும். இவை நிகண்டுகளுக்கும், அகரமுதலிகளுக்கும் முன்னோடியாகக் கருதப்படுகின்றன. நிகண்டுகளில் ஒருபொருட் பல்பெயர், ஒருசொற் பல்பொருள், தொகைப்பெயர் என்னும் மூன்று பெரும் பிரிவுகள் உண்டு. பிற்கால நிகண்டுகள் ஒரு பொருளைத் தரும் பலசொற்களையும் திரட்டித் தந்தன. திவாகர நிகண்டு என்னும் நிகண்டு நூல் கிபி 8 ஆம் நூற்றாண்டில் வாழ்ந்த திவாகர முனிவர் என்பவரால் இயற்றப்பட்டது. ஒத்த கருத்துடைய சொற்களை ஒரிடத்தில் சேர்த்துக் கொடுப்பது, தெஸாரஸ் (thesauras) ஆகும். திவாகர நிகண்டின் முதல் பத்துத் தொகுதிகளும் இவ்வமைப்பைக் கொண்டவை. சொற்களுக்குப் பொருள் கூறும் அகராதி நூல் கி.பி. 16 ஆம் நூற்றாண்டில் வாழ்ந்த இரேவண சித்தர் என்பவரால் இயற்றப்பட்டது. இதுவே அகரவரிசையில் அமைந்த முதல் அகராதி. இவை யாவும் செய்யுள் வடிவில் அமைந்தவை.

ஆங்கிலச் சொல்வலை

1985 இல் உளவியல் பேராசிரியரான ஜார்ஜ் ஆர்மிடேஜ் மில்லரின் தலைமையில் பிரின்ஸ்டன் பல்கலைக்கழக உணர்வு அறிவியல் ஆய்வகத்தில் முதல் ஆங்கிலச் சொல்வலை உருவாக்கப்பட்டது.

நவம்பர் 2012 இல் ஆங்கிலச் சொல்வலை தரவுத்தளத்தில் மொத்தம் 155 ஆயிரம் சொற்கள் 175 ஆயிரம் நேர்ச்சொல் தொகுப்பில் 207 ஆயிரம் சொல்லுக்கான பொருள் இணைகளாக உள்ளன.

தமிழ்ச் சொல்வலை

நேர்ச்சொல் தொகுப்புகள் மற்ற நேர்ச்சொல் தொகுப்புகளுடன் உறவுகளை முறைப்படுத்தி கருத்துருக்களின் அடுக்குவரிசை அமைப்பை உருவாக்குகின்றன. எடுத்துக்காட்டாக, ‘உயிரினம்’ போன்ற மிகவும் பொதுவான வகையிலிருந்து ஆரம்பித்து ‘விலங்கு’ போன்ற இடைப்பட்ட வகை, பின்னர் ‘எருமை மாடு’ போன்ற மிகவும் குறிப்பிட்ட வகை வரை.

பல உட்பிரிவுகளை உள்ளடக்கியப் பெரும்பிரிவுகள் (hypernyms) மிகவும் பொதுவானவை. துல்லியமாகக் குறிக்கப்படும் உட்பிரிவுகள் (hyponyms) மிகவும் குறிப்பிட்டவை. எருமை மாடு விலங்குகளின் ஒரு வகையாக இருக்கிறது என்று கூறலாம். இதை “ஒரு வகை (is a type of)” உறவு என்று கூறுகிறோம். Meronym என்பது மற்றொன்றின் ஒரு பகுதியைக் குறிக்கும் ஒரு சொல். மூக்கு (meronym) என்பது முகத்தில் “ஒரு பகுதி (is a part of)”. Holonym என்பது ஒரு முழுமையைக் குறிக்கும் சொல். வேறு சொற்கள் அதன் பகுதிகளைக் குறிக்கும். இந்த சிலை (holonym) களிமண்ணால் “செய்யப்பட்டது (is made of)”. சுருக்கமாகச் சொல்லப்போனால் சொல்வலை என்பது சொற்பொருள் உறவுகளால் ஒன்றாக இணைக்கப்பட்ட தமிழ் சொற்களின் தரவுத்தளமாகும். இது வரைபட அமைப்பிலுள்ள (graph structure) மிகையூட்டம் செய்யப்பட்ட அகராதி மற்றும் நிகண்டு போன்றது.

ஒத்த பொருட்களை அடிப்படையாகக் கொண்ட சொற்களை ஒன்றாகத் தொகுப்பதால் சொல்வலை ஒரு நிகண்டுவை (thesaurus) மேலோட்டமாக ஒத்திருக்கிறது. எனினும், சில முக்கியமான வேறுபாடுகள் உள்ளன. முதலாவதாக, சொல்வலையிலுள்ள சொற்களின் எழுத்துச் சரங்களை மட்டுமல்ல, சொற்களின் குறிப்பிட்ட பொருள் உணர்வையும் இணைக்கின்றன. இதன் விளைவாக, பிணையத்தில் ஒன்றோடு ஒன்று நெருங்கிய தொடர்பில் காணப்படும் சொற்கள் சொற்பொருளியல் ரீதியாக ஐயமகற்றப்படுகின்றன. இரண்டாவதாக, சொற்களில் சொற்பொருள் உறவுகளை சொல்வலை அடையாளப்படுத்துகிறது. ஆனால் நிகண்டுவில் உள்ள சொற்களின் தொகுப்புகள் பொருள் ஒற்றுமை தவிர வேறெந்த வெளிப்படையான வடிவத்தையும் பின்பற்றுவதில்லை.

சொல்வலையின் உறவுகளின் பெரும்பகுதி, ஒரே சொல்வகையிலுள்ள (POS) சொற்களை இணைக்கிறது. இவ்வாறு, சொல்வலையில் பெயர்ச்சொற்கள், வினைச்சொற்கள், பெயரடை மற்றும் வினையடை ஒவ்வொன்றுக்கும் உண்மையில் நான்கு துணை வலைகள் உள்ளன.

[மும்பையிலுள்ள இந்திய தொழில்நுட்பக் பயிலகம் 18 இந்திய மொழிகளுக்கான சொல்வலைகளைத் தயாரித்திருக்கிறது.](#) தமிழ்ச் சொல்வலையில் 16,312 பெயர்ச்சொற்கள், 2803 வினைச்சொற்கள், 5827 பெயரடைகள் மற்றும் 477 வினையடைகள் உள்ளன. இந்தி சொல்வலை மற்றும் அதன் நிரலாக்க இடைமுகம் குனு பொது உரிமம் (GPL) 3.0 மற்றும் சொற்களஞ்சியம் (Lexicon) குனு கட்டற்ற ஆவணங்கள் (FDL) உரிமம் கீழ் வெளியிடப்பட்டுள்ளன. மற்ற மொழிகளும் இதே உரிமங்களில்தான் இருக்க வேண்டும். இவற்றை பதிவிறக்கம் செய்ய [இந்தப் படிவத்தை நிரப்பிக் கோரிக்கை அனுப்ப வேண்டும்.](#)

இதன் இணையதளத்தில் ‘படி’ என்ற சொல்லைத் தேடினால் கீழ்க்கண்ட விவரம் கிடைக்கிறது:

Number of Synset for “படி” : 12

Synset ID : 5954

POS : verb

Synonyms : படி, வாசி, உச்சரி,

Gloss : புத்தகம், கவிதை, கதை கடிதம் போன்றவற்றை உச்சரிப்பது

Example statement : “மோஹித் தன் அப்பாவின் கடித்தத்தை படித்துக் கொண்டிருந்தான்”

Gloss in English : interpret something that is written or printed; “read the advertisement”; “Have you read Salman Rushdie?”

Other Synsets: பயணப்படி, குளத்தின் படி, கற்றுக்கொள், பாலில் ஆடை படிந்துள்ளது, தானியத்தை அளக்கும் பாத்திரம், சட்டையில் கறை படிந்திருக்கிறது முதலியன.

சொல்லின் பொருளில் ஐயமகற்றல் (Word Sense Disambiguation – WSD)

இயந்திர மொழிபெயர்ப்பு, தகவல் மீட்பு, வினாவிற்கு விடை அளிப்பது போன்ற பயன்பாடுகளுக்கு சொல்லின் பொருளில் ஐயமகற்றல் மிகவும் முக்கியத்துவம் வாய்ந்தது. எடுத்துக்காட்டாக, தமிழில் இருந்து ஆங்கிலத்துக்கு இயந்திர மொழிபெயர்ப்பு செய்யும்போது ‘ஆறு’ என்ற சொல் வந்தால் முதலில் அது ‘நீர்வழி’ என்ற பொருளில் வந்ததா அல்லது ‘ஒரு எண்’ என்ற பொருளில் வந்ததா என்று தீர்மானிக்க வேண்டும். அதன் பின்னரே அதற்குச் சரியான ஆங்கிலச் சொல் தேர்ந்தெடுக்க முடியும்.

சொல்லின் பொருளில் ஐயமகற்றலுக்கு 1986 ஆம் ஆண்டில் மைக்கேல் லெஸ்க் (Michael Lesk) அறிமுகப்படுத்திய வினைச்சரம் முக்கியமானது. ஒரு சொல்லின் அக்கம்பக்கத்துச் சொற்கள் அதே பொதுவான தலைப்பைப் பகிர்ந்து கொள்கின்றன என்ற கருத்தை இந்த வினைச்சரம் அடிப்படையாகக் கொண்டது. இயல்மொழி கருவித்தொகுதி (Natural Language ToolKit – NLTK) என்பது பைதான் மொழியில் இயல்மொழி ஆய்வுக்கான கருவித்தொகுதி. இது சொல்லின் பொருளில் ஐயமகற்றலுக்கு சொல்வலை பயன்படுத்தி செம்மையான லெஸ்க் வழிமுறையை செயல்படுத்துகிறது. ஒரு தெளிவற்ற சொல்லையும் அதன் சூழல் வாக்கியத்தையும் கொடுத்தால், சூழல் வாக்கியத்திற்கும் நேர்ச்சொல் தொகுப்பின் சொற்பொருள் விளக்கத்துக்கும் இடையில் மிக அதிகமான சொற்களைப் பொதுவாகக் கொண்ட ஒரு நேர்ச்சொல் தொகுப்பை லெஸ்க் வினைச்சரம் தேர்வு செய்து தரும்.

22. அடிச்சொல், தண்டுச்சொல் மற்றும் சொற்பகுப்பாய்வு

பகுதி விகுதி இடைநிலை சாரியை

சந்தி விகாரம் ஆறினும் ஏற்பவை

முன்னிப் புணர்ப்ப முடியும் எப்பதங்களும்

நன்னூல். பதவியல் – 133 (13 ஆம் நூற்றாண்டில் பவணந்தி முனிவரால் எழுதப்பட்ட தமிழ் இலக்கண நூல்)

சொற்பகுப்பியல் (morphology)

சொற்கள் எப்படி சிறிய அலகுகளால் உருவாக்கப்படுகின்றன என்ற சொல் கட்டமைப்பு ஆய்வை சொற்பகுப்பியல் அல்லது உருபனியல் என்று சொல்கிறோம். சொல் என்பது என்ன? மொழியை எழுதும்போது, சொல் என்பது இரு இடைவெளிகளுக்கு இடையே இருப்பது என்று நாம் சொல்லலாம். ஆனால் சொல் என்பது ஒரு பொருளுள்ள கருத்து. “மைசூர் பாகு” என்பது ஒரு பொருளுள்ள கருத்து. மொழியைப் பேசும்போது இதை நாம் ஒரே சொல்லாகச் சேர்த்து உச்சரிக்கிறோம். ஒரு சொல்லைப் பகுக்கும்போது வரும் யாவற்றினும் மிகச்சிறிய அடிப்படை அலகை உருபன் (morpheme) என்கிறோம். இது அடிச்சொல், ஒட்டுகள் இரண்டுக்குமே பொதுவான பெயர்.

நாம் ஒரு சொல்லைப் பற்றித் தேடும்போது, தேடல் பெட்டியில் உள்ளிட்டது மட்டுமல்லாமல் அதன் பிற சாத்தியமான வடிவங்களுக்கும் பொருத்தமான முடிவுகளைக் கண்டுபிடிக்கத்தான் விரும்புகிறோம். எடுத்துக்காட்டாக “சிங்கப்பூர்” என்று தேடல் பெட்டியில் உள்ளிடுகிறோம் என்று வைத்துக் கொள்வோம். நமக்கு “சிங்கப்பூரின்”, “சிங்கப்பூருடன்”, “சிங்கப்பூரிலேயே” என்ற சொற்கள் இருக்கும் பக்கங்களும் தேவைதானே?

சொற்பகுப்பியல்படி வளம் மிகுந்த தமிழில் சொல் வடிவம் அதிகமாக மாறும். ஒரே அடிச்சொல்லை வைத்து காலம், திணை, இடம், எண், பால் போன்ற இலக்கண வகைகளுக்குத் தகுந்தவாறு ஒட்டுகள் சேரும். இம்மாதிரி ஒட்டுகள் சேர்வது வினைச்சொல்லில் மிக அதிகம், அடுத்து பெயர்ச்சொல்.

பகுபதம் என்பது பகுக்க அல்லது பிரிக்கக்கூடிய வகையில் அமைந்த சொல். மேலே கண்ட நன்னூல் வரிகளப்படி பகுபதத்தைப் பகுதி, விகுதி, இடைநிலை, சாரியை, சந்தி, விகாரம் என்ற ஆறாகப் பகுக்க முடியும் என்று தமிழ் பகுபத உறுப்பு இலக்கணம் சொல்கிறது.

அடிச்சொல் பிரிப்பி (Lemmatizer or Lemma extractor)

ஒரு ஆவணத்தில் முக்கிய சொற்களைப் பட்டியலிடவேண்டும் என்று வைத்துக் கொள்வோம். எடுத்துக்காட்டாக, ஒரு ஆவணத்தில் ‘மின்னூல்’, ‘மின்னூல்கள்’, ‘மின்னூலை’, ‘மின்னூலின்’ என்று சொல்லின் வேறுபாடுகள் இருக்கலாம். இவற்றை முழுச்சொற்கள் என்று சொல்லலாம். எந்த சொல் அடிக்கடி வருகிறது என்று புள்ளிவிவரம் பார்க்கும்போது இவற்றை தனித்தனியே எண்ணிக்கை எடுத்தால் சரியாக வருமா? இவ்வெல்லாச் சொற்களையும் அடிச்சொல் ஆக்கிக் கூட்டினால், இது மின்னூல் பற்றிய ஆவணம் என்று உடன் கூறிவிடலாம் அல்லவா? இதற்கு நல்ல அடிச்சொல் பிரிப்பி தேவை (Lemmatizer or Lemma extractor). முழுச்சொல்லை பகுபத உறுப்பிலக்கணம்படி பிரித்தெடுத்து கடைசியில் வரும் அடிச்சொல் சொல்லகராதியில் உள்ளதாக வெளியிட வேண்டும். இது கடினமான வேலை.

ஆங்கில தண்டுச்சொல் பிரிப்பி (Stemmer)

சில வேலைகளுக்கு இவ்வளவு மெனக்கெடாமல் ஒட்டுகளை உத்தேசமாக வெட்டிவிட்டுத் தண்டுச் சொல்லை வைத்துக் கொள்ளலாம். இது சொல்லகராதியில் உள்ள சொல்லாக இருக்க வேண்டிய

அவசியமில்லை. இதைத் தண்டுச்சொல் பிரிப்பி என்று சொல்கிறோம். எடுத்துக்காட்டாக, ஆங்கிலத்தில் பரவலாகப் பயன்படுத்தப்படும் போர்ட்டர் தண்டுச்சொல் பிரிப்பி (Porter stemmer) 'apple', 'apples' இரண்டையுமே 'appl' என்ற தண்டுச்சொல்லாகப் பிரிக்கிறது. 'appl' என்பது அகராதி சொல் அல்ல. இருப்பினும் நாம் தேடும் சொல்லையும் ஆவணத்திலுள்ள சுட்டுச் சொல்லையும் அதே தண்டுச் சொல்லாக்குவதால் ஒப்பிடும்போது வேலைக்கு ஆகிறது.

இரு சொற்களும் அதே லத்தீன் வேரிலிருந்து பெறப்பட்டாலும், நாம் 'universe', 'university' என்ற சொற்களை ஒப்பிட விரும்பமாட்டோம்.

வினவல்களுக்கும் சுட்டு ஆவணங்களுக்கும் தண்டுச்சொல் பிரிப்பியைப் பயன்படுத்தினால், சொற்களின் வேறுபட்ட வடிவங்களை ஒன்றாக்கி தேடல் திறனை அதிகரிக்க முடியும். ஆனால் வினவல், ஆவணங்கள் இரண்டுக்குமே ஒரே தண்டுச்சொல் பிரிப்பியைப் பயன்படுத்துவது மிக முக்கியம்.

தமிழ் தண்டுச்சொல் பிரிப்பி

[திரு. தாமோதரன் ராஜலிங்கம் ஒரு தண்டுச் சொல் பிரிப்பி உருவாக்கி திறந்த மூலமாகப் பகிர்ந்து கொண்டுள்ளார்.](#) இது விதிமுறைப்படி படிப்படியாக ஒட்டுகளை அகற்றக்கூடியது. தமிழ்ச் சொற்களில் இது முன்னொட்டுகளையும் பின்னொட்டுகளையும் அகற்றுகிறது. ஆனால் கூட்டுச் சொற்களுக்கு (எடுத்துக்காட்டு: கருப்பு + குதிரை = கருங்குதிரை) வேலை செய்யாது.

இதை [ஸ்னோபால் \(Snowball\) மொழியைப்](#) பயன்படுத்தி செயல்படுத்தியுள்ளார். ஸ்னோபால் என்பது தகவல் மீட்புக்குத் தேவையான வினைச்சரங்களை உருவாக்கும் ஒரு சிறிய சரம் செயலாக்க மொழி ஆகும்.

இதற்கு மாறாக, அடிச்சொல் பிரிப்பியோ சொல்லின் அடிப்படை புரியாமல் மேலெழுந்த வாரியாக தலையையும் வாலையும் வெட்டுவது இல்லை. அதற்குப் பதிலாக சொற்களின் சரியான அடிப்படை வடிவங்களைப் பெற சொல்வலை (wordnet) போன்ற ஒரு தரவுத்தளத்தை நம்பியுள்ளது. சில அடிச்சொல் பிரிப்பிகளுக்கு சொல்வலைக் குறியீடையும் (POS) உள்ளிட வேண்டும். இல்லையெனில், அது சொல்லை ஒரு பெயர்ச்சொல் என்றே கருதுகிறது. இறுதியாக, அடிச்சொல் பிரிப்பியால் தெரியாத சொற்களைக் கையாள முடியாது.

ஆங்கிலத்திலும் மற்ற மொழிகளிலும் சொற்பகுப்பாய்வு (morphological analysis)

அபெர்டியம் (Apertium) என்பது இயந்திர மொழிபெயர்ப்புக்குத் திறந்த மூல மென்பொருள். இந்த திட்டத்தின் ஒரு பகுதியான மொழித் தொழில்நுட்பம் கருவிப்பெட்டி (Language Technology Toolbox) ஆங்கிலத்துக்கும் மற்ற சில ஐரோப்பிய மொழிகளுக்கும் உருபனியல் ஆய்வு செய்ய உதவுகிறது. அதாவது கொடுத்த சொல்லை அதன் அடிச்சொல்லாகவும், ஒட்டுகளாகவும் பிரிக்கிறது. மேலும் அதன் எதிர் செயல்முறையையும் செய்யும். அதாவது அடிச்சொல்லையும், ஒட்டுகளையும் சேர்த்து முழுச்சொல் வடிவத்தையும் உண்டாக்கும்.

தமிழில் உருபனியல் ஆய்வு

அடிச்சொல் மற்றும் ஒட்டுகளைச் சேர்த்துப் பொதுவாக உருபன் (morpheme) என்று சொல்கிறோம். அடிச்சொல் (base morpheme) தனித்து செயல்படும், அகராதியில் உள்ள, சொல் (எ.கா., வாங்கு). ஒட்டுகள் (bound morpheme) அகராதியில் இருக்காது (எ.கா., னார்). இவை அடிச்சொல்லுடன் முன்னொட்டாகவோ அல்லது பின்னொட்டாகவோதான் வர முடியும்.

உருவமைப்புப்படி ஒரு சொல்லின் எல்லா உருபன்களையும் பிரித்தெடுப்பதை உருபனியல் ஆய்வு (morphological analysis) என்று சொல்கிறோம். [அண்ணா பல்கலைக்கழகத்தில் கேபிசி ஆராய்ச்சி மையம் தமிழ் உருபனியல் பகுப்பாய்வி \(Morphological Analyser\) ஒன்றைத் தயாரித்துள்ளது.](#) இதை 3 மில்லியன் சொற்கள் கொண்ட மைய இந்திய மொழிகள் கழகத்தின் உரைத்தொகுப்பை வைத்து சோதனை செய்தபோது 95% சொற்களை ஆய்வு செய்ய முடிந்ததாம். இது முன்னுதாரணம் சார்ந்த

அணுகுமுறையில் (paradigm-based approach) வரையறுக்கப்பட்ட நிலை இயந்திரமாக (Finite State Machine) செயல்படுத்தப்பட்டுள்ளதாம்.

மும்பை இந்திய தொழில்நுட்பக் கழகத்தில் அனுப் குருசுக்குட்டன் இந்திய மொழிகளுக்கான இயல்மொழி நிரலகம் ஒன்றை திறந்த மூலமாக பொது உரிமம் 3.0 கீழ் வெளியிட்டுள்ளார். இதில் ஒரு தமிழ் உருபனியல் பகுப்பாய்வியும் உள்ளது. ஒரு சொல்லைக் கொடுத்தால் அதைப் பகுப்பாய்வு செய்து உருபன்களாகத் தரும். ஆக்க நிலை ஒட்டுகளையும் (Derivational suffixes) உட்பிணைப்பு ஒட்டுகளையும். (Inflectional suffixes) அடையாளம் கண்டு விடுமாம்.

ஆக்க நிலை ஒட்டுகள் சொல்வகையை மாற்றும். எடுத்துக்காட்டாக ஆங்கிலத்தில் 'person' என்பது பெயர்ச்சொல். 'al' என்று ஒட்டு சேர்த்தால் 'personal' என்ற பெயரடை. தமிழில் 'ஒட்டு' என்ற வினைச்சொல்லுடன் 'நர்' என்ற ஒட்டு சேர்த்தால் 'ஒட்டுநர்' என்ற பெயர்ச்சொல் கிடைக்கும்.

23. சார்புநிலை பிரிப்பி, சுட்டுப்பெயர் தீர்வு, தலைப்பு பிரித்தெடுத்தல்

சார்புநிலை பிரிப்பி

இயல்மொழியைப் புரிந்து கொள்வது கடினமானது! “I saw a girl with a telescope” என்ற வாக்கியத்தைப் பாருங்கள். தொலைநோக்கி வைத்திருந்த பெண்ணை நீங்கள் பார்த்தீர்களா அல்லது நீங்கள் தொலைநோக்கி மூலம் பார்த்தீர்களா? இது ஆங்கில எடுத்துக்காட்டு. எனினும் தமிழிலும் இதே பிரச்சினை உள்ளது. “ஜெயலலிதாவைத் திருப்திப்படுத்தும் ஆர்வத்தில் சட்டசபையில் காங்கிரஸ் உறுப்பினர்கள் பேச ஆரம்பித்ததும் அ.தி.மு.கவினர் குறுக்கீடு செய்தனர்.” என்று செய்தித் தலைப்பு. ஜெயலலிதாவைத் திருப்திப்படுத்தும் ஆர்வம் காங்கிரஸுக்கா அல்லது அ.தி.மு.கவினருக்கா? இம்மாதிரி வாக்கியக் கட்டமைப்பில் தெளிவின்மைக்கு முறையான வழியில் தீர்வு காண வாக்கியத்தை “பாகுபடுத்தல் (Parsing)” செய்ய வேண்டும்.



பாகுபடுத்தியில் இரண்டு வகைகள் உள்ளன. சொற்றொடர் அமைப்பு (Phrase structure) பாகுபடுத்தி சொற்றொடர்களை அடையாளம் காணுவதில் கவனம் செலுத்துகிறது. சார்புநிலை (Dependency) பாகுபடுத்தி சொற்களிடையே உறவுகளில் கவனம் செலுத்துகிறது. சார்புநிலை பாகுபடுத்தியில் இரண்டு விதங்கள். முதலாவதில் சொற்களுக்கு இடையில் உறவைக் குறிக்க வகை விவரச் சீட்டு போடுவோம். இரண்டாவதில் வகை விவரச் சீட்டு கிடையாது. எந்த சொற்கள் மற்ற எந்த சொற்களைச் சார்ந்துள்ளன என்று கோடு போட்டு மட்டும் காட்டுவோம். பாகுபடுத்தியின் வெளியீடு பொதுவாக CoNLL (Conference on Natural Language Learning) கோப்பு வடிவத்தில் இருக்கும். XML வடிவத்திலும் வெளியிடப்படலாம்.

இவற்றில் சார்பு நிலை ஆய்வுதான் தமிழ் போன்ற கட்டற்ற சொல் வரிசை கொண்ட மொழிகளுக்குத் தகுந்தது. வினைச் சொல்லைக் கட்டமைப்பின் மையத்தில் வைத்து மற்ற சொற்கள் அதன்மீது நேரடியாகவோ அல்லது மறைமுகமாகவோ சார்ந்து இருப்பதை ஆய்வு செய்கிறோம்.

கிளைப்பட வங்கி

சொல்வகைக் குறியீடு செய்த தொகுப்பை எடுத்து அதற்கு அடுத்த படியாக சார்புநிலை ஆய்வு செய்து கிளைப்பட வங்கி தயாரிக்கிறோம். இதை ஓரளவு தானியங்கியாகத் தயாரிக்கலாம். முதலில் சார்புநிலை பிரிப்பியை ஓட்டி வெளியீடு செய்யலாம். பின்னர் மொழியியலாளர்கள் இதை சரி பார்க்கலாம். இவ்வாறு ஆய்வு செய்து குறியிட்ட உரைத்தொகுப்பை கிளைப்பட வங்கி (treebank) என்று சொல்கிறோம்.

இயந்திரக் கற்றல் மூலம் பாகுபடுத்தி தயார் செய்ய இம்மாதிரி கிளைப்பட வங்கிகள் அவசியம் தேவை. உலகமொழிகளில் ஒரு சில மொழிகளுக்குத் தான் இம்மாதிரி கிளைப்பட வங்கிகள் தயார் செய்யப்பட்டுள்ளன. இந்திய மொழிகளில் இந்திக்கு மட்டுமே கிளைப்பட வங்கி உள்ளது. உலகத்தில் அனேகமாக மற்ற எல்லா மொழிகளுக்கும் இம்மாதிரி கிளைப்பட வங்கிகள் திறந்த மூலமாகவோ அல்லது ஆய்வுக்கு மட்டும் இலவசமாகவோ கிடைக்கின்றன. ஆனால் இந்திக்குக் கிடையாது. அவர்களுடைய இணையதளத்திற்குச் சென்று பார்த்தால் முதலில் பதிவு செய்யச் சொல்கிறார்கள்.

தெலுங்கு மொழிக்கு ஒரு மில்லியன் சொற்கள் கொண்ட கிளைப்பட வங்கி தயாரித்துக் கொண்டிருக்கிறார்களாம்.

[செக் குடியரசு நாட்டில் சார்லஸ் பல்கலைக்கழகத்தில் லோகநாதன் ராமசாமி 600 வாக்கியங்கள் மட்டுமே கொண்ட தமிழ் கிளைப்பட வங்கி தயாரித்திருக்கிறார். TrEd என்பது கிளைப்படம் போன்ற கட்டமைப்புகளைப் பார்க்கவும், மாற்றங்கள் செய்யவும் உதவும் திறந்த மூல மென்பொருள். குனு பொது உரிமம் \(GPL\) கீழ் கிடைக்கிறது. இதைப் பதிவிறக்கி நிறுவினால் மேற்கண்ட கோப்பைத் திறந்து கிளைப்படங்களைப் பார்க்கவும், விரும்பினால் மாற்றங்கள் செய்யவும் முடியும்.](#)

சார்புநிலை பிரிப்பி திறந்த மூல பைதான் நிரல்

2014 இல் அமெரிக்கா கார்னகி மெல்லான் பல்கலைக்கழகத்தில் பிரதீப் பிரபாகர் ரவீந்திரன் தமிழுக்கு சார்புநிலை பிரிப்பிக்கான பைதான் நிரலை திறந்த மூலமாக கிட்ஹப்பில் வெளியிட்டுள்ளார். இந்த நிரலால் அடைய முடிந்த மிகச்சிறந்த அடையாளமிடாத சார்பு நிலை இணைப்பு மதிப்பெண் (unlabeled dependency attachment score) ஒரு உரையில் 69.7% மற்றொரு உரையில் 67.9%. “இந்தத் திட்டத்தில் ஒரு பெரிய பின்னடைவு என்னவென்றால் நல்ல தரமான சொல் பிரிப்பி மற்றும் உருபனியல் பகுப்பாய்வுக் கருவிகளின் பற்றாக்குறைதான். ‘அம்’, ‘ஆக’ போன்ற பல்வேறு இடைச்சொற்களைப் பிரிக்க சொல் பிரிப்பிகள் இருந்தால் நிச்சயமாக தரத்தை மேம்படுத்த முடியும்.” என்று இவர் சொல்கிறார். சார்புநிலை குறியீடு செய்யும் நிரலுடன் சேர்த்து கிளைப்படங்களாக குறியீடு செய்த ஒரு உரைத்தொகுப்பும் வெளியீடு செய்யப்போவதாகச் சொன்னார்கள். ஆனால் அது பின்னர் வெளிவந்ததா என்று தெரியவில்லை.

[HamleDT \(HArmonized Multi-LanguagE Dependency Treebank\) என்ற திட்டத்தின் கீழ் பல மொழிகளின் ஒத்திசைவான சார்புநிலை கிளைப்பட வங்கிகள், பொதுவான குறியீடு பாணியில் தொகுத்துள்ளார்கள். 29 மொழிகளின் கிளைப்படங்களை ஆய்வு செய்து, வேறுபாடுகள் பட்டியல் தயாரித்துள்ளார்கள். இதில் தமிழும் ஒன்று. CoNLL வடிவத்திலும் XML வடிவத்திலும் பதிவிறக்கம் செய்யலாம்.](#)

சுட்டுப்பெயர் தீர்வு

சுட்டுப்பெயர் தீர்வு (Coreference resolution or Anaphora resolution) என்பது ஒரு உரையில் உள்ள அதே உருபொருளைக் குறிக்கும் அனைத்துச் சொற்றொடர்களையும் கண்டுபிடிக்கும் பணியாகும். ஆவணச் சுருக்கம், கேள்வி பதில் மற்றும் தகவல் பிரித்தெடுத்தல் போன்ற உயர் நிலை இயல் மொழிப் புரிதல் பணிகளுக்கு இது ஒரு முக்கியமான படியாகும்.

இதையே வேறு விதமாகச் சொல்லலாம். ஒரு உரையைப் பற்றிய சரியான விளக்கம் பெற, அதிலுள்ள சுட்டுப்பெயர்களையும் மற்றும் பிற குறிப்பிடும் சொற்களையும் சரியான நபர்களுடனோ அல்லது பொருளுடனோ இணைக்க வேண்டும். அந்த உரையில் குறிப்பிடப்பட்டுள்ள பல்வேறு நபர்களின் மற்றும் பொருள்களின் முக்கியத்துவத்தை மதிப்பிடவும் இதுவே தேவைப்படுகிறது. எடுத்துக்காட்டாக, ஒரு உரையில் “அப்துல் கலாம்” மற்றும் “அவர்கள்” என்று இருந்தால் அவை இரண்டும் அதே நபரைத்தான் குறிப்பிடுகின்றனவா என்று தீர்வு செய்ய வேண்டும்.

John Smith, CFO of Prime Corp since 1986,
saw his pay jump 20% to \$1.3 million
as the 57-year-old also became
the financial services co.'s president.

சுட்டுப்பெயர் தீர்வு எடுத்துக்காட்டு

[ஸ்டான்போர்ட் CoreNLP திறந்த மூல இயல்மொழிக் கருவிகள் தொகுப்பில்](#) CorefAnnotator என்ற ஒரு சுட்டுப்பெயர் தீர்வுக் கருவி உள்ளது. மூன்று வகையான சுட்டுப்பெயர் தீர்வுக் கருவிகள் உள்ளன. ஒன்று விதிமுறை சார்ந்தது, வேகமாக ஓடும். இரண்டாவது புள்ளிவிவரம் சார்ந்தது. மூன்றாவது நரம்பியல் அடிப்படையிலானது, மிகவும் துல்லியமானது ஆனால் மெதுவாக ஓடும். [இதைப் பயிற்றுவிக்கும் நிரல் பைதானில் எழுதியது](#). எந்த மொழிக்கும் பழக்கலாம் என்று சொல்கிறார்கள். பயிற்றுவிக்க தரவு CoNLL வடிவத்தில் இருக்க வேண்டும். வரைபடச் செயலகம் (GPU) சில்லு உள்ள கணினி தேவை. இதுவரை ஆங்கிலம் மற்றும் சீன மொழிகளுக்கு மட்டுமே பயிற்சி அளிக்கப்பட்டுள்ளது. ஆகவே தமிழுக்குப் பழக்க முடியுமா என்று தெரியவில்லை.

தலைப்பு பிரித்தெடுத்தல்

ஒரு ஆவணம் எதைப் பற்றியது என்று தெரிய வேண்டுமானால் அதைப் படித்துத் தான் பார்க்க வேண்டும். நீங்கள் ஒரு பத்திரிகை அலுவலகத்தில் வேலை பார்க்கிறீர்கள் என்று வைத்துக் கொள்வோம். பல ஆண்டுகளாக வெளியிடப்பட்ட பல்லாயிரக்கணக்கான கட்டுரைகளை ஏழெட்டு தலைப்புகளில் பிரிக்கச் சொல்கிறார்கள். கணினியைப் பயன்படுத்தி தலைப்பு பிரித்தெடுப்பது (Topic Extraction or Modeling) இந்த வேலையை எளிதாக்கும். இம்மாதிரி வேலையைப் பொதுவாக உரை சுரங்கவேலை (text mining) என்றும் சொல்கிறோம்.

ஒரு குறிப்பிட்ட தலைப்பைப் பற்றிய ஆவணத்தில் அந்தத் தலைப்பு பற்றிய சொற்கள் அதிகமாக இருக்கும் என்று உள்ளுணர்வாகச் சொல்லலாம். எடுத்துக்காட்டாக, ஒரு ஆவணத்தில் 'மருத்துவர்', 'நோயாளி', 'மருத்துவமனை' போன்ற சொற்கள் அடிக்கடி வந்தால் அதன் தலைப்பு 'நலம்' என்றும் 'பண்ணை', 'பயிர்', 'நெல்' போன்ற சொற்கள் அடிக்கடி வந்தால் அதன் தலைப்பு 'விவசாயம்' என்றும் கூறலாம்.

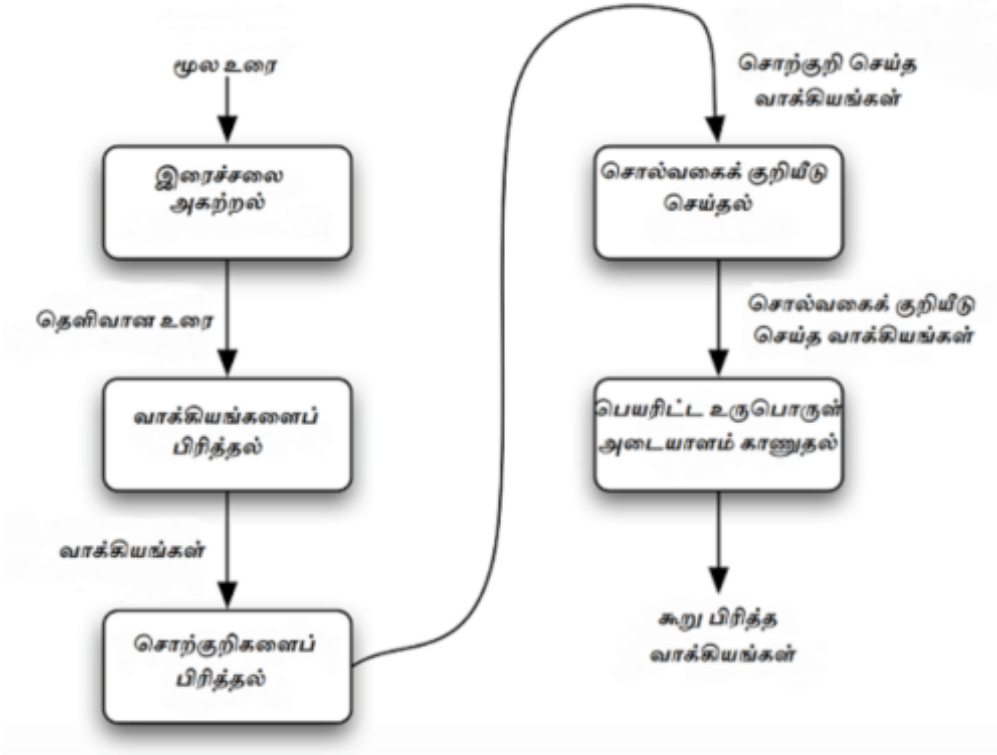
[மாசுசுசெட்ஸ் பல்கலையில் மேலட் \(MALLET\) திறந்த மூலத் திட்டம்](#) புள்ளிவிவர மற்றும் இயந்திரக் கற்றல் அடிப்படையிலான ஆவண வகைப்பாடு, தொகுப்பு, தலைப்பு மாதிரியாக்கம், தகவல் பிரித்தெடுத்தல் மற்றும் பிற இயல்மொழி கருவிகளை வெளியீடு செய்துள்ளது. இது ஜாவா நிரலாக்க மொழியில் எழுதப்பட்டது. இயக்கத்திலுள்ள இந்தத் திட்டத்துக்கு 30 நிரலாளர்கள் பங்களிக்கிறார்கள். இந்த திட்டத்தை விரும்பி 626 பயனர்கள் விண்மீன் குறியீடு அளித்துள்ளனர். சொற்குறிகளைப் பிரித்து உள்ளீடு செய்துவிட்டால் இது எந்த மொழிக்கும் வேலை செய்யும் என்று சொல்கிறார்கள்.

24. இயல்மொழி ஆய்வு கருவித் தொடரி

இயல்மொழி ஆய்வில் எந்தவொரு வேலையை நிறைவேற்றவும் பல பணிகளை ஒன்றன்பின் ஒன்றாகச் செயல்படுத்த வேண்டும். எந்தவொரு சிறிய வேலைக்குக் கூட பெரும்பாலும் கீழ்க்கண்ட பணிகள் இன்றியமையாதவை:

- வாக்கியங்களைப் பிரித்தல்
- சொற்களைப் (நிறுத்தற் குறிகளையும் சேர்த்து) பிரித்தல்
- சொல்வகைக் குறியீடு செய்தல்
- அடிச்சொல்லையோ, தண்டுச்சொல்லையோ பிரித்தெடுத்தல்

இதன் பின்னர், தேவையைப் பொருத்து, சார்புநிலைப் பிரிப்பியை வைத்து கிளைப்பட வங்கிகளாகவோ அல்லது சொற்பகுப்பியல் ஆய்வியை வைத்து உருபங்களாகவோ பிரிப்போம். ஆக, குறைந்த பட்சம் ஐந்தாறு பணிகளை ஒன்றன்பின் ஒன்றாகச் செயல்படுத்துவோம்.



நிரல்கூறுகளாக இயல்மொழி ஆய்வு கருவித் தொடரி

இவ்வாறு தொடர்ச்சியாகச் செய்யும் பணிகளுக்கு மென்பொருள் ஒரே திரளாக இல்லாமல் தனித்தனியாக பராமரிக்கக் கூடிய நிரல்கூறுகளாக (modular) இருப்பது நல்லது. மேலும், பிரச்சினை வந்தால் ஒவ்வொரு நிரல்கூற்றின் உள்ளீட்டையும், வெளியீட்டையும் தனித்தனியாக நம்மால் ஆய்வு செய்ய இயல வேண்டும். இதைத்தான் இயல்மொழி ஆய்வு கருவித் தொடரி (tool chain) அல்லது குழாய்த்தொடர் (pipeline) என்று சொல்கிறோம்.

அதாவது, இயல்மொழி ஆய்வில் ஒரு குழாய்த் தொடர் என்பது தற்சார்புடைய நிரல்கூறுகளின் சங்கிலி ஆகும். ஒவ்வொன்றும் அதற்கு முன்னதாக உள்ள நிரல்கூற்றின் வெளியீட்டை எடுத்து தன்னுடைய செய்முறையை செய்து அடுத்த நிரல்கூறுக்கு அனுப்புகிறது.

எடுத்துக்காட்டாக, பைதானில் எழுதப்பட்ட இயல்மொழி கருவித் தொகுப்பில் (NLTK) அடிப்படைக் கருவிகள் உள்ளன. இந்தக் கருவிகளை வைத்து உங்கள் வேலைக்குத் தகுந்தவாறு நீங்கள் ஒரு கருவித் தொடரியை அமைத்துக் கொள்ளலாம். ஏனெனில் ஒவ்வொரு நிரல் கூறும் வெவ்வேறு வழிமுறைகளைப் பயன்படுத்தி நிறைவேற்றப்படலாம், மேலும் கருவித் தொடரியின் மற்ற பகுதிகளை பாதிக்காமல் ஒரு வினைச்சரத்தை மாற்றலாம்.

தற்பொழுது கிடைக்கும் இயல்மொழி ஆய்வுக் கருவிகள்

நாம் முன்பு கூறியது போல, தமிழுக்கான கருவிகள் பெரும்பாலும் கிடைப்பதில்லை. செயல் விளக்கத்துக்காக சில கருவிகள் இணையதளத்தில் செயலியாக வெளியிடப்பட்டுள்ளன. பொதுவாக இவற்றில் ஒரு சொல்லை உள்ளிட்டு, வேலை செய்கிறதா என்று பார்க்க முடியும், அவ்வளவுதான். வேறு சில கருவிகள் பதிவிறக்கமாக அல்லது எழுதிக் கேட்டால் கிடைக்கக்கூடும். அப்படிக் கிடைத்தாலும் இவை வரைபடப் பயனர் இடைமுகம் கொண்டவை. ஒரு நேரத்தில் சிறிய அளவிலான சொற்களை மட்டுமே உள்ளிட முடியும். இவற்றையெல்லாம் விட மிக முக்கியமாக, தனித்தனியாகத்தான் வேலை செய்யும். நாம் மேலே விவரித்தபடி கருவித் தொடரியாக இணைக்க முடியாது.

இரைச்சலை அகற்றல் (Noise removal)

நீங்கள் பலவீனமான ஒலியளவு கொண்ட வானொலி அலைவரிசையைக் கேட்க முயல்கிறீர்கள் என்று வைத்துக்கொள்வோம். இரைச்சலைக் குறைத்தால்தான் சமிக்ஞை தெளிவாக வந்து சேரும். தேவையான ஒலியை நன்றாகக் கேட்க முடியும். அதே போலவே உரையிலுள்ள உட்பொருளைப் பிரித்தெடுக்க இரைச்சலைக் குறைக்க வேண்டும்.

இரைச்சல் அகற்றுவதற்கான ஒரு பொது அணுகுமுறை இரைச்சல் நிறைந்த சொற்குறிகளின் ஒரு பட்டியலைத் தயார் செய்வது. பின்னர் இந்த சொற்குறிகளை ஒன்றொன்றாக உரை முழுவதும் தேடி நீக்குதல். இணையத்திலும், சமூக ஊடகங்களிலும் எடுத்த உரைகளில் இணையத்தள முகவரிகள் (URLs), கொத்துக்குறிகள் (hashtags), வேற்று மொழிச் சொற்கள் போன்றவை இருக்கலாம். அஃகுப்பெயர்களுக்கும் (acronyms) சுருக்கங்களுக்கும் பதிலாக அவற்றின் முழு வடிவங்களை மாற்றி வைக்க வேண்டும். இம்மாதிரி சிறப்பு வடிவங்களைக் கையாளும் போது மற்றொரு அணுகுமுறை தேடுகுறித்தொடர்களைப் (regex) பயன்படுத்துவது.

நாம் அடிக்கடி பயன்படுத்தும் இடைச்சொற்களில் (Stopwords) பொருள் குறைவாக இருப்பதால் இவை ஒரு உரையை மற்ற உரைகளிலிருந்து வேறுபடுத்த உதவுவதில்லை. எடுத்துக்காட்டாக ஒரு, என்று, மற்றும், இந்த, இது, என்ற, கொண்டு, என்பது, பல, ஆகும்... போன்ற சொற்கள். நாம் சில நேரங்களில், மேலும் செயலாக்கும் முன், உரையிலிருந்து இவற்றை வடிகட்டி எடுத்துவிட வேண்டியுள்ளது. [நான் இந்த 125 சொற்கள் கொண்ட பட்டியலை TSM சொற்களஞ்சியத்தில் முதல் 300 சொற்குறிகளிலிருந்து \(சொற்களும் நிறுத்தற்குறிகளும்\) தயாரித்தேன்.](#)

நாம் ஒரு உரையை ஆய்வு செய்யும் போது, வேறெந்த வேலையும் செய்யும் முன்னர் இவற்றைச் செய்ய வேண்டும். ஆகவே இவற்றைச் செயல்படுத்தல் முன்னோடிகள் (pre-processing) என்றும் சொல்கிறோம்.

வாக்கியங்களைப் பிரித்தல் (Sentence segmentation)

வாக்கியங்களின் எல்லைகளைக் கண்டறிதல், அதாவது பத்திகளாக உள்ள உரையிலிருந்து வாக்கியங்களாகப் பிரிக்கும் செயல். இந்தப் பணியை மேலோட்டமாகப் பார்த்தால் மிக எளிதாகத் தெரிகிறது. ஆனால் நாம் புள்ளியை வாக்கிய முடிவைக் காட்ட நிறுத்தற் குறியாகவும் பயன்படுத்துகிறோம், சுருக்கங்களுக்கும், பதின்ம எண்களுக்கும் பயன்படுத்துகிறோம். ஆகவே புள்ளி வந்தால் வாக்கியம் முடிந்துவிட்டது என்று கூற முடியாது.

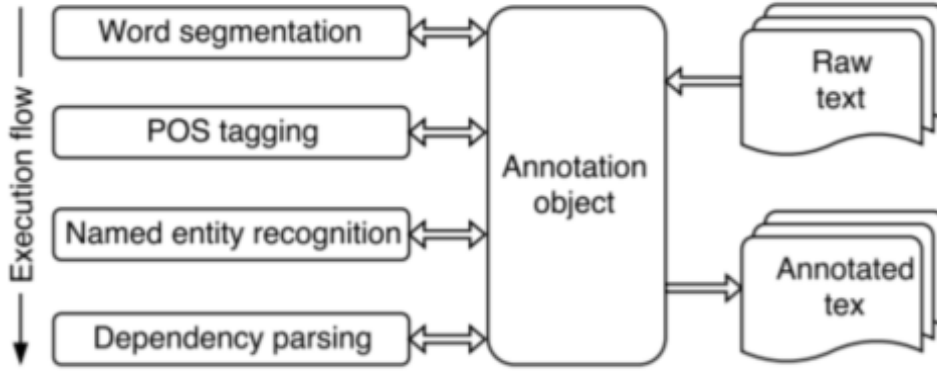
ஆங்கில வாக்கியப் பிரிப்பான்களால் தமிழின் தனிப்பட்ட அம்சங்களை சரியாகக் கையாள இயலவில்லை. எ.கா., பெயர் முதலெழுத்துகள், முன்னொட்டுகள், சுருக்கக் குறியீடுகள் ஆகியவை. ஆகவே இதற்காக [நான் NLTK Punkt tokenizer-ஐத் தமிழுக்குத் தயார் செய்துள்ளேன்.](#) இது இயந்திரக் கற்றல் (machine learning) மூலம் சாதாரண சொல்வங்கி (raw text corpus) வைத்து செய்யப்பட்டது.

சொற்களையும் நிறுத்தற் குறிகளையும் பிரித்தல் (Tokenization)

ஒரு வாக்கியத்தில் சொற்களும் நிறுத்தற் குறிகளும் இருக்கும். இவற்றைப் பொதுவாக சொற்குறிகள் (Tokens) என்று சொல்கிறோம். இடைவெளிகளை வைத்து இவற்றைப் பிரிக்கலாம். ஆனால் முற்றுப் புள்ளி முக்கால் புள்ளிகளுக்கு முன் இடைவெளி இல்லாமலும் இருக்கலாம். 'ச.ம.உ.' என்பதைப் புள்ளிகளில் பிரிக்காமல், ஒரே சொல்லாக எடுக்க வேண்டும்.

மேலும் இடைவெளி இருப்பதால் “தென் சென்னை” இரண்டு சொற்களா அல்லது இரண்டும் சேர்ந்து ஒரு இடத்தைக் குறிப்பதால் ஒரே சொல்லா? ‘சட்ட மன்ற உறுப்பினர்’ என்பதையும் இடைவெளிகளை வைத்துத் தனித்தனி சொற்களாகப் பிரிக்காமல், ஒரே சொல்லாக எடுக்க வேண்டும். ஆனால் இவற்றை சொற்குறிகள் பிரிக்கும்போது செய்ய இயலாவிட்டால், பின்னர் பெயரிட்ட உருபொருள் பிரித்தெடுக்கும்போது செய்யலாம்.

இதன் பின்னர், செய்யும் வேலையைப் பொருத்து, சொல்வகைக் குறியீடு செய்தல், அடிச்சொல்லையோ, தண்டுச்சொல்லையோ பிரித்தெடுத்தல் போன்ற செயல்முறைகளைக் கருவித் தொடரியில் இணைக்கலாம்.



வியட்நாமிய இயல்மொழி ஆய்வு குழாய்த்தொடர்

நடைமுறை எடுத்துக்காட்டு: வியட்நாமிய இயல்மொழி ஆய்வு குழாய்த்தொடர்

சொற்களைப் பிரித்தெடுத்தல், சொல்வகைக் குறியீடு, பெயரிட்ட உருபொருள் அடையாளம் காணுதல் மற்றும் சார்புநிலை பிரிப்பி ஆகிய முக்கியப் பணிகளில் பயனர்களுக்கு உதவுவதற்கும், அடுத்து வரும் செய்முறைகளின் ஆய்வாளர்கள் மற்றும் கருவி நிரலாளர்களுக்கு ஆதரவாக இருக்கவும் ஆங்கிலத்தில் ஸ்டான்ஃபோர்ட் CoreNLP கருவித்தொகுப்பு போன்ற ஒரு இயல்மொழி குழாய்த்தொடரை வியட்நாமிய மொழிக்கு உருவாக்குவதற்கான தேவை உள்ளது என்று முடிவு செய்தனர். [வியட்நாமிய மொழிக்கு இயல்மொழி ஆய்வுக் குறியீடு செய்யும் இந்த VnCoreNLP என்ற குழாய்த்தொடர் திட்டத்தைத் தொடங்கினர்.](#) இது ஜாவா (Java) மொழியில் எழுதப்பட்டுள்ளது.

[இந்த நிரல் தொகுப்பை படைப்பாக்கப் பொதுமங்கள் அனுமதியில் பகிர்ந்துள்ளனர்.](#) வணிகம் அல்லாத வேலைகளுக்கு இலவசமாகப் பயன்படுத்தலாம். வியட்நாமிய மொழிக்கு இதுவரை உள்ள கருவிகளைவிட மிகவும் துல்லியமாக வேலை செய்கிறது மற்றும் வேகமாக உள்ளது. எனவே இதை பெரிய அளவிலான தரவுகளைக் கையாளுவதற்குப் பயன்படுத்தலாம். எளிதாகப் பயன்படுத்த முடியும். பயனர்கள் வெளிப்புற சார்புகளை நிறுவ வேண்டியதில்லை. கட்டளை வரி அல்லது ஜாவா (Java) பயன்பாட்டு நிரலாக்க இடைமுகத்திலிருந்து (API) இயக்க முடியும்.

நடைமுறை எடுத்துக்காட்டு: தாய்லாந்து இயல்மொழி ஆய்வு கருவித்தொகுப்பு

NLTK என்பது இயல்மொழி ஆய்வுக்கு பைதான் மொழியில் எழுதப்பட்டது. இது ஒருங்குறியை ஆதரிப்பதால் தமிழுக்கும் ஓரளவு பயன்படுத்த முடியும். இதிலுள்ள தேடுகறித்தொடர் நிரலகம்

ஒருங்குறியை ஆதரிக்கவில்லை. ஆகவே தமிழில் பெயரிட்ட உருபொருள் அடையாளம் காணப் பயன்படாது. இந்த NLTK போல [TLTK என்று தாய்லாந்து இயல்மொழி ஆய்வு கருவித்தொகுப்பு தயாரித்துள்ளார்கள்](#). இதில் வாக்கியத்திலிருந்து சொற்களைப் பிரித்தெடுத்தல், சொல்வகைக் குறியீடு செய்தல், பெயரிட்ட உருபொருள் அடையாளம் காணுதல் போன்ற பல வேலைகளுக்கு கருவிகள் உள்ளன.

25. தமிழ் – ஆங்கிலம் இயந்திர மொழிபெயர்ப்பு

இயந்திர மொழிபெயர்ப்புக்கு மூன்று வகையான அணுகல்கள் உள்ளன. இவை விதி சார்ந்த இயந்திர மொழிபெயர்ப்பு (Rule-Based Machine Translation – RBMT), புள்ளிவிவர இயந்திர மொழிபெயர்ப்பு (Statistical Machine Translation – SMT) மற்றும் கலப்பு (Hybrid) இயந்திர மொழிபெயர்ப்பு.

விதி சார்ந்த இயந்திர மொழிபெயர்ப்பு

விதி சார்ந்த இயந்திர மொழிபெயர்ப்பில் இந்த இரண்டு வகைகள் உண்டு. மேலோட்டமான மொழிமாற்றம் (Shallow transfer) மற்றும் ஆழ்ந்த மொழிமாற்றம் (Deep transfer). அபெர்டியம் (Apertium) ஒரு கட்டற்ற திறந்த மூல விதி சார்ந்த இயந்திர மொழிபெயர்ப்பு தளம் ஆகும். இது குனு பொது உரிமத்தின் கீழ் வெளியிடப்பட்டுள்ளது. இது ஒரு மேலோட்டமான-மொழிமாற்ற இயந்திர மொழிபெயர்ப்பு அமைப்பு ஆகும். தற்பொழுது, அபெர்டியம், இந்தி – உருது சேர்த்து, 40 நிலையான மொழி இணைகளை வெளியிட்டுள்ளது.

இது இந்தி – உருது போன்ற நெருக்கமாக தொடர்புடைய மொழிகளுக்கு இடையில் மொழிபெயர்க்கவே வடிவமைக்கப்பட்டது. இருப்பினும் இது சமீபத்தில் மாறுபட்ட மொழி இணைகளுக்கும் விரிவாக்கப்பட்டுள்ளது. ஒரு புதிய இயந்திர மொழிபெயர்ப்பு அமைப்பு உருவாக்க, XML வடிவங்களில் மொழி தொடர்பான தரவுகளை (அகராதிகள், விதிகள்) மட்டும் உருவாக்க வேண்டும்.

தமிழ் ஒரு கட்டற்ற சொல் வரிசை மொழி. ஆகவே தமிழிலிருந்து ஆங்கிலத்துக்கு மொழிபெயர்ப்பதில் துல்லியம் குறைவு. ஆங்கிலத்திலிருந்து தமிழுக்கு மொழிபெயர்ப்பதில் துல்லியம் அதிகம்.

மூல மொழியில் ஒரு சொற்பகுப்பியல் ஆய்வியும் இலக்கு மொழியில் ஒரு சொற்பகுப்பியல் உருவாக்கியும் அபெர்டியம் இயந்திர மொழிபெயர்ப்பை தமிழுக்கு செயற்படுத்த முக்கியமானவை. [ஹைதராபாத் பல்கலையில் பரமேஸ்வரி இவற்றை உருவாக்க செய்த முயற்சியின் முடிவுகள் இங்கே](#). சுமார் 68 ஆயிரம் சொற்கள் கொண்ட அகராதியைப் பயன்படுத்தி இருவேறு உரைத்தொகுப்புகளில் 85% சொற்களை மொழிபெயர்ப்பு செய்ய முடிந்ததாம்.

இயந்திர மொழிபெயர்ப்பு செய்முறை வரிசை

விதி சார்ந்த இயந்திர மொழிபெயர்ப்புக்கு கீழ்க்கண்ட செய்முறை வரிசையைப் பின்பற்றுகிறோம்:

1. பகுப்பாய்வு: பகுப்பாய்வின் போது, உள்ளீட்டைப் பாகுபடுத்த மூல மொழியில் செயலிகளும், இலக்கணம் மற்றும் அகராதியும் தேவை.
2. மொழிமாற்றம்:
 1. வாக்கியக் கட்டமைப்பு மாற்றத்திற்கு ஒரு இலக்கணம் தேவை.
 2. சொற்களை மொழிமாற்றம் செய்ய இருமொழி அகராதி தேவை.
3. உருவாக்கம்: கடைசியாக, இலக்கு மொழியில் உருவாக்க இலக்கணமும், அகராதியும் இறுதி மொழிபெயர்ப்புக்குத் தேவை.



தமிழ் – ஆங்கிலம் இயந்திர மொழிபெயர்ப்பு

புள்ளிவிவர இயந்திர மொழிபெயர்ப்பு

மருத்துவ, நிதி அல்லது தொழில்நுட்பம் போன்ற குறிப்பிட்ட துறைக்கான பயிற்சி தரவுகளைப் பயன்படுத்தி பயிற்றுவித்தால், புள்ளிவிவர இயந்திர மொழிபெயர்ப்பு இயந்திரம் உயர் வெளியீடு தரத்தை நிரூபித்துள்ளது.

ஒரு குறிப்பிட்ட துறைக்கு குறைந்த பட்சம் இரண்டு மில்லியன் சொற்கள் தேவை என்று பொதுவாகக் கருதப்பட்டாலும், அதைவிடக் குறைந்த தரவுகளை வைத்தே ஏற்றுக்கொள்ளத்தக்க தரத்தை அடைய முடியும் என்று சொல்கிறார்கள். நல்ல தரமான சீரமைக்கப்பட்ட இருமொழித் தொகுப்புகள் உருவாக்க செலவு அதிகம். எனினும் உருவாக்கிய பின்னர் அது நிறுவனத்துக்கு மதிப்பு மிக்க சொத்தாகும்.

புள்ளியியல் இயந்திர மொழிபெயர்ப்புக்கு [மோசஸ் \(Moses\) ஒரு திறந்த மூல மென்பொருள்](#). ஒரு மூல மொழியிலிருந்து ஒரு இலக்கு மொழிக்கு உரைகளை மொழிபெயர்ப்பு செய்ய புள்ளிவிவர மாதிரிகளை பயிற்றுவிக்க பயன்படுகிறது. பயிற்றுவித்த இந்த மாதிரிகளை வைத்து மூல மொழி உரைகளை மொழிபெயர்க்கலாம். பயிற்சி கொடுக்க இரண்டு மொழிகளில் வாக்கியங்கள் சீரமைக்கப்பட்ட ஒரு இணை உரைத்தொகுப்பு வேண்டும்.

இது LGPL உரிமத்தின் கீழ் வெளியீடு செய்யப்பட்டுள்ளது. விண்டோஸ், மேக் மற்றும் லினக்ஸ் கணினிகளில் ஓடும். ஐரோப்பிய ஒன்றிய நிதியுதவியில் உருவாக்கப்பட்டது.

கலப்பு இயந்திர மொழிபெயர்ப்பு

கலப்பு (Hybrid) இயந்திர மொழிபெயர்ப்பில் பொதுவாக RBMT மற்றும் SMT இரண்டும் உண்டு. இதை SYSTRAN போன்ற தனியுரிம நிறுவனங்களே பெரும்பாலும் பயன்படுத்துகின்றன.

நரம்பியல் இயந்திர மொழிபெயர்ப்பு

நரம்பியல் இயந்திர மொழிபெயர்ப்பு (Neural Machine Translation – NMT) என்பது ஒரு பெரிய செயற்கை நரம்பியல் பிணையத்தைப் பயன்படுத்தி அடுத்து வரும் சொற்களின் தொடர்ச்சியைக் கணிக்கிறது. இது பொதுவாக ஒருங்கிணைந்த முன்மாதிரியை வைத்து முழு வாக்கியங்களை உருவாக்க முயற்சிக்கிறது. நரம்பியல் இயந்திர மொழிபெயர்ப்பு என்பது புள்ளிவிவர இயந்திர மொழிபெயர்ப்பில் (SMT) வழக்கமாகச் செய்வதைவிடத் தலைகீழான மாற்றம் அல்ல. ஆகவே இதுவும் புள்ளிவிவர இயந்திர மொழிபெயர்ப்பில் ஒரு வகைதான். ஆனால் மிகவும் மேம்படுத்தப்பட்ட வகை.

இதில் மொழி மாதிரியும், மொழிபெயர்ப்பு மாதிரியும், மறு சீரமைப்பு மாதிரியும் தனித்தனியாகக் கிடையாது. ஆனால் அடுத்து வரும் ஒவ்வொரு சொல்லாகக் கணிக்கும் ஒரு தனி வரிசைமுறை மாதிரி மட்டும் உண்டு. முழு மூல வாக்கியத்தையும் மற்றும் ஏற்கனவே தயாரிக்கப்பட்ட இலக்கு சொல் வரிசைமுறையையும் வைத்து இந்த வரிசைக் கணிப்பு நிர்ணயிக்கப்படுகிறது. ஆழமான NMT இதனுடைய நீட்டிப்பு ஆகும். ஒரே ஒரு அடுக்குக்குப் பதிலாக இது பல நரம்பியல் பிணைப்பு அடுக்குகளை செயல்படுத்துகிறது. இதை ஆழ்ந்த கற்றல் (Deep Learning) என்றும் சொல்கிறார்கள்.

இயந்திர மொழிபெயர்ப்பில் நரம்பியல் பிணையங்களைப் பயன்படுத்துவதற்கான முதல் அறிவியல் அறிக்கை 2014 இல் வெளிவந்தது. முதன்முதலாக 2015 ஆம் ஆண்டில், பொது இயந்திர மொழிபெயர்ப்பு போட்டியில் NMT அமைப்பு இருந்தது. அடுத்த ஆண்டு வெற்றியாளர்களில் 90% NMT பயன்படுத்தியவர்கள். 2016 க்குள், சிறந்த இயந்திர மொழிபெயர்ப்பு அமைப்புகளில் பெரும்பாலானவை NMT அமைப்புகள்தான். கூகிள், மைக்ரோசாப்ட் மற்றும் யாண்டெக்ஸ் மொழிபெயர்ப்பு சேவைகள் இப்போது NMT ஐயே பயன்படுத்துகின்றன.

[திறந்த மூல நரம்பியல் இயந்திர மொழிபெயர்ப்பு அமைப்பு, OpenNMT, ஹார்வர்ட் இயல்மொழி ஆய்வுக் குழுவால் வெளியிடப்பட்டது.](#) OpenNMT என்பது நரம்பியல் இயந்திர மொழிபெயர்ப்புக்கான திறந்த மூல முன்முயற்சியாகும். இது அண்மைய ஆழ்ந்த கற்றல் தொழில்நுட்பத்தையே பயன்படுத்துகிறது. இது தற்போது 3 முக்கிய செயலாக்கங்கள் கொண்டது: லுவா (Lua), பைதான் – பைடார்ச்(PyTorch), பைதான் – டென்சார்ஃப்ளோ(TensorFlow). இவை மூன்றுமே தற்போது பராமரிக்கப்படுகின்றன. [மூல நிரல்கள் கிட்ஹப்பில் பகிரப்பட்டுள்ளன.](#) OpenNMT MIT உரிமத்தின் கீழ் வெளியிடப்பட்டுள்ளது.

[பங்களிப்போர் எண்ணிக்கையையும், பயனர் ஆர்வத்தைக் குறிக்கும் விண்மீன் குறியீடுகளையும் பார்த்தால் கூகிளின் டென்சார்ஃப்ளோதான் \(TensorFlow\) சவாலில்லாமல் முதலிடத்திலுள்ளது.](#) இது அபாச்சி 2.0 திறந்த மூல உரிமத்தின் கீழ் வெளியிடப்பட்டுள்ளது. இது பைதான் மொழியில் எழுதப்பட்டுள்ளது. ஆனால் கருநிரல் C ++ மற்றும் CUDA வில் எழுதப்பட்டுள்ளது. CUDA என்பது வரைபடச் செயலகம் (GPU) நிரலாக்கம் செய்ய என்விடியா (Nvidia) மொழி.

OpenNMT மென்பொருளை பயன்படுத்தி பயிற்றுவிக்க உங்களுக்கு இரண்டு கோப்புகள்தான் தேவை – மூல மொழிக்கோப்பு மற்றும் இலக்குமொழிக்கோப்பு. ஒவ்வொரு வரியிலும் ஒரு வாக்கியம்தான் இருக்க வேண்டும். சொற்களுக்கு இடையில் இடைவெளி இருக்க வேண்டும்.

வரைபடச் செயலகம் (GPU) சில்லு உள்ள கணினி தேவை

OpenNMT நிரலைப் பயன்படுத்தி பயிற்றுவிக்க NVIDIA நிறுவனத்தால் தயாரிக்கப்பட்ட வரைபடச் செயலாக்கச் (GPU) சில்லு உள்ள கணினி தேவை. அதுவும் CUDA என்ற கருவித்தொகுதி மென்பொருள் ஓடக்கூடியதாக இருக்க வேண்டும். [உங்கள் கணினியில் இவை உள்ளனவா என்று பார்க்க இங்கே செல்லவும்.](#) பெரிய அளவில் தரவைச் செயலாக்க வேண்டியிருப்பதால், பொதுவாக எந்த இயந்திர கற்றல் திட்டத்திற்கும் வரைபடச் செயலகம் சில்லு வைத்த கணினி தேவைப்படுகிறது.

இயந்திர மொழிபெயர்ப்பு அணுகல்களில் நல்லதும் கெட்டதும்

RBMT அமைப்புகள் மொழியின் மூன்று முக்கியமான குறைபாடுகளை எதிர்கொள்கின்றன:

- 1) ஒரு இயந்திரத்தைப் பொருத்தவரை மொழியின் உள்ளார்ந்த தெளிவினமை. மனிதர்களைப் போல சொல்லின் பொருள் சார்ந்த உள்ளுணர்வு இயந்திரங்களுக்குக் கிடையாது.
- 2) விலக்கு அடிப்படையிலான இலக்கண அமைப்பு.
- 3) எப்போதும் விரிவாகிக்கொண்டிருக்கும் கலைச்சொற்கள் அகராதி.

NMT அணுகுமுறை மொழி கட்டமைப்பை நன்றாகக் கற்றுக் கொள்ளும், ஆனால் RBMT அல்லது SMT போல் நீண்ட சொல் பட்டியலை ஞாபகம் வைத்துக் கொள்வதில் அவ்வளவு சிறப்பாக இல்லை. சொற்களின் பட்டியலை RBMT நன்றாகக் கையாளும், ஆனால் வாக்கிய அமைப்பு சரியாக வராது. SMT இவற்றுக்கு இடையில் உள்ளது. வாக்கியங்கள் சரளமாக வருவது போலிருக்கும், ஆனால் சில நேரங்களில் மொழிபெயர்ப்பு முற்றிலும் தவறானதாக இருக்கலாம்.

NMT பற்றிய சுவாரசியமான அம்சம் என்னவென்றால், PBSMT (சொற்றொடரை அடிப்படையாகக் கொண்ட SMT) யை விட இரைச்சல் மிகுந்த தரவுக்கு சகிப்புத்தன்மை அதிகம். பொதுவான NMT அமைப்புகளில் ஒரு குறிப்பிடத்தக்க பலவீனம் என்னவென்றால் மிகவும் அரிதான வார்த்தைகளை அவை சரியாக மொழிபெயர்ப்பதில்லை.

26. சொற்பிழைத் திருத்தி

தமிழுக்குச் சொல்திருத்தியே தேவையில்லை என்றொரு கருத்து

ஆங்கிலத்தில் உச்சரிப்பை வைத்து எழுத்துக்கோர்வை சொல்ல முடியாது. ஆகவே எழுத்துப்பிழைகள் நிகழ்வதற்கான சாத்தியங்கள் உள்ளன. தமிழிலோ எப்படி உச்சரிப்போ அப்படியே எழுதுகிறோம் (Phonetic language). ஆகவே தமிழுக்குச் சொல்திருத்தியே தேவையில்லை என்றொரு கருத்து நிலவுகிறது.

ஆனால் மயங்கொலி என்று சொல்லப்படும் ல-ள-ழ, ண-ந-ன, ர-ற ஆகியவற்றில் எது சரி என்று தெரியாமல் நாம் தவறு செய்கிறோம். மேலும் 'fat finger' என்று ஆங்கிலத்தில் கூறப்படும் தவறான விசையை அழுத்துவதால் ஏற்படும் தட்டச்சுப் பிழைகளையும் பார்க்கிறோம். ஒற்றுப்பிழை எனும் வல்லினம் மிகும் மிகா இடங்கள் எது என்பதில் பெரும்பாலோருக்கு ஐயமுண்டு. மற்றும் எச்சொற்களைச் சேர்த்து எழுதுவது எவற்றைப் பிரித்து எழுதுவது என்பதிலும் குழப்பம் அதிகம். விக்கிமூலம் போன்ற இடங்களில் ஒளி எழுத்துணரி பிழைகளும் நிறைய உள்ளன. ஆகவே தமிழுக்குச் சொல்திருத்தி தேவையதான். ஆனால் ஆங்கிலச் சொல்திருத்தியின் நகலாக இல்லாமல் தனித்தன்மை வாய்ந்ததாக இருக்க வேண்டும். எனினும் பாமா விஜயம் திரைப்படத்தில் வருவது போல 'முடித்துத் தருவோம்' என்பதற்குப் பதிலாக 'முடி திருத்துவோம்' என்று எவரும் தட்டச்சு செய்தால், இவையும் சொற்தொகுதியில் இருக்கும் சொற்கள் என்பதால், பிழையென்று எந்தச் சொல்திருத்தியாலும் சுட்டிக்காட்ட முடியாது!

உரையில் சொற்பிழை சரிபார்ப்பு பின்வரும் படிகளைக் கொண்டுள்ளது

- உரையிலிருந்து நாம் சரிபார்க்க விரும்பும் சொற்களைப் பிரித்தெடுத்துக் கொள்ளுதல் (tokenizing).
- ஒரு சொல் அகராதியில் உள்ளதா என்று முதலில் பார்த்தல்.
- இல்லாவிட்டால், அடுத்து அடிச்சொல் (lemma) மற்றும் ஒட்டுகளை விதிப்படி சேர்த்து அந்தச் சொல் வருகிறதா என்று பார்த்தல்.
- இரண்டிலும் இல்லாவிட்டால், அது தவறான சொல். தவறான சொற்களுக்கு, அகராதியிலும், ஒட்டுகளைச் சேர்த்தும் இதே போன்ற வேட்பு சொற்களைக் (candidate words) கண்டறிவதன் மூலம் திருத்தங்களைப் பரிந்துரைத்தல்.

வேட்பு சொற்கள்

ஒரு சொல்லைப் பிழை என்று அடையாளமிட்டவுடன் திருத்தங்களைப் பரிந்துரைக்க நாம் வேட்பு சொற்கள் பட்டியலைத் தயாரிக்க வேண்டும். இந்தப் பட்டியலைத் தயாரிக்க நாம் திருத்தும் தூரம் (edit distance) என்ற கருத்துருவைப் பயன்படுத்துகிறோம். ஆங்கிலத்தில் ஒரு திருத்துத் தூரத்தில் உள்ள எழுத்துப்பிழைகளின் வகைகள் இவை:

1. செருகல் பிழை (insertion): ஒரு கூடுதல் எழுத்து சேர்தல், எ.கா., truly -> truey. ஒரு முக்கியமான சிறப்பு வழக்கு மீண்டும் அதே எழுத்து, எ.கா., occasion -> occassion. (இந்தப் பிழை தமிழில் அரிது.)
2. விடுபடு பிழை (deletion): ஒரு எழுத்து விட்டுப்போய் விட்டது, எ.கா., unnecessary -> unnecesary
3. பதிலீட்டுப் பிழை (substitution): ஒரு எழுத்துக்குப் பதிலாக வேறொரு எழுத்து உள்ளது, எ.கா., separate -> sepearate.
4. இடமாற்றுப் பிழை (transposition): இரண்டு தொடர்ச்சியான எழுத்துகள் இடமாற்றம் ஆகியுள்ளன, எ.கா., receive -> recieve.

பெரும்பாலான சொற்பிழைகள் இந்த ஒரு திருத்துத் தூரத்தில்தான் வரும்.



சொற்பிழைத் திருத்தி

ஹன்ஸ்பெல் திறந்த மூல சொற்பிழைத் திருத்தி

ஆங்கிலத்தில் முன்னொட்டுகளும் பின்னொட்டுகளும் உள்ளன. இருப்பினும் தமிழில் வினைச்சொற்களுக்கும் பெயர்ச்சொற்களுக்கும் இருப்பது போல அந்த அளவுக்குக் கிடையாது. ஆகவே ஆங்கில சொல்திருத்திகளைத் தமிழுக்குத் தகுந்தவாறு மாற்றி அமைக்க முடியாது.

ஹன்ஸ்பெல் (Hunspell) திறந்த மூல மென்பொருள் முதலில் ஹங்கேரிய மொழிக்காக எழுதப்பட்டது, இதன் முக்கிய பங்களிப்பாளர் லாஸ்லோ நெமெத். ஹங்கேரிய மொழியும் பல ஒட்டுகள் கொண்டது. எ.கா., *ház-a-i-é-i-nek* (அவர்களின் வீடுகளுடைய). ஹன்ஸ்பெல் சொற்பிழைத் திருத்தியில் இம்மாதிரி சொற்கள் முற்றிலும் ஒட்டுகளைப் பிரிப்பதன் மூலம் தீர்க்கப்படும். இது தமிழ் போன்ற சொற்றொடரியல் ரீதியாக வளமான மொழிகளுக்கும் ஒத்து வரும். இதைத் தமிழுக்குப் பயன்படுத்த வேண்டுமானால் அடிச்சொற்களைக் கொண்ட அகராதியையும் முன்னொட்டு, பின்னொட்டுகளுக்கான விதிகளையும் தயாரிக்க வேண்டும்.

தமிழுக்கு நான் உருவாக்கிய ஹன்ஸ்பெல் பின்னொட்டு விதிகள்

இதற்கான கோப்புகளையும், விவரங்களையும் இங்கே காணலாம். சொல் ஒட்டுகளுக்கான விதிகளை .aff கோப்பு கொண்டிருக்கிறது. எந்தெந்த ஒட்டுகள் ஒரு அடிச்சொல்லுடன் சேர்ந்து மாற்று சொற்களை உருவாக்கும் என்ற விதிகளை இக்கோப்பில் எழுதியுள்ளேன்.

- வினைச்சொல்: இதற்கான பின்னொட்டு விதிகள் ஒவ்வொரு அடிச்சொல்லுக்கும் சுமார் 8500 சொல் வேறுபாடுகளை உருவாக்குகின்றன. 1000-க்கும் மேலான வினைச்சொற்களைக் குறியிட்டுள்ளேன்.
- பெயர்ச்சொல்: இதற்கான பின்னொட்டு மற்றும் முன்னொட்டு விதிகள் சேர்ந்து ஒவ்வொரு அடிச்சொல்லுக்கும் சுமார் 2500 சொல் வேறுபாடுகளை உருவாக்குகின்றன. 4000-க்கும் மேலான பெயர்ச்சொற்களைக் குறியிட்டுள்ளேன்.

என் சொற்பட்டியலில் சுமார் 330,000 சொற்கள் மட்டுமே உள்ளன. இருப்பினும் இந்த சுமார் 30,000 வரிகள் கொண்ட ஒட்டு விதிகள் மொத்தம் 20 மில்லியனுக்கும் மேலான வினைச்சொல் மற்றும் பெயர்ச்சொல் வேறுபாடுகளை உருவாக்குகின்றன!

ஹன்ஸ்பெல்லில் முக்கியக் குறைபாடு என்னவென்றால், ஒற்றுப்பிழை கண்டுபிடிக்க நேரடி ஆதரவு இல்லை. தவிர இது சொற்களைச் சேர்த்து எழுதுவது பிரித்து எழுதுவதால் ஏற்படும் பிழைகளையும் கையாளுவதில்லை. இது திறந்த மூல நிரல் ஆகையால் இதை நீட்சி செய்ய முடியுமா என்று பார்க்க

வேண்டும். இது C++ இல் எழுதப்பட்டது. ஆனால் பைதான் மற்றும் ஜாவா உள்ளிட்ட பல மொழிகளுக்கு இடைமுகங்கள் மற்றும் மாற்றியெழுதிய நிரல்கள் உள்ளன.

லேங்குவேஜ் டூல் திறந்த மூல சொற்பிழைத் திருத்தி

[திறந்த மூல சொற்பிழைத் திருத்தி பற்றி மலேசியா நண்பர்கள் சி. ம. இளந்தமிழ் & வே. இளஞ்செழியன் எழுதிய கட்டுரை இங்கே.](#) இக்கட்டுரையில் லேங்குவேஜ் டூல் (Language Tool) பற்றிய முக்கிய குறிப்புகள் சில. “லேங்குவேஜ் டூல் ஒற்றுப்பிழை மட்டுமின்றி சொற்களைச் சேர்த்தும் பிரித்தும் எழுதுவதால் ஏற்படும் தவறுகளையும் கண்டுபிடிக்கும் வசதியைக் கொண்டிருக்கிறது.

இது ஜாவா நிரலாக்க மொழியைக் கொண்டு எழுதப்பட்ட ஒரு மென்பொருளாகும் . இதனை டேனியல் நேபர் என்ற ஒரு ஜெர்மானிய நிரலாளர் உருவாக்கினார் . இது கடந்த பத்து ஆண்டுகளாகத் திறவூற்றுச் சமூகத்தால் தொடர்ந்து மேம்படுத்தப்பட்டு இன்று 20 க்கும் மேற்பட்ட மொழிகளை ஆதரிக்கிறது.

இத்திருத்தி இலக்கணப்பிழைகளைக் கண்டுபிடிக்க இரண்டு அடிப்படைக் கோப்புகளைப் பயன்படுத்துகின்றது. முதலாவது கோப்பு – tamil.dict – சொற்பட்டியையும் சொற்பகுதிக் குறிப்பையும் (POS tags) கொண்டிருக்கிறது. இரண்டாவது கோப்பு – grammar.xml – இலக்கண விதிகளைக் கொண்டிருக்கிறது. இக்கோப்புகளைப் பயன்படுத்தி ஒரு சொற்றொடரில் உள்ள பிழைப் பாங்குகளை (error patterns) கண்டுபிடித்து, அவற்றைப் பிழையெனக் குறிப்பிடுவதோடு, அப்பாங்குகளை எவ்வாறு திருத்தி அமைக்கலாம் எனவும் லேங்குவேஜ் டூல் பரிந்துரைக்கிறது.”

[லேங்குவேஜ் டூல்](#) இணையதளத்தில் தமிழைத் தேர்வு செய்து உரையை உள்ளிட்டு சரிபார்த்தால் பின்வருமாறு சொற்களைச் சேர்த்தும் பிரித்தும் எழுதப் பரிந்துரை செய்கிறது. “வேண்டும்” என்ற வினைமுற்று வடிவம்முன் இடம்விட வேண்டும்: “கற்க வேண்டும்”. பெயரெச்சத்தின் பின் ‘போது’ வந்தால் சேர்த்து எழுத வேண்டும்: “தேவைப்படும்போது”.

ஆக, தமிழில் சொற்பிழைகளைக் கண்டுபிடிக்க நம்மிடம் ஹன்ஸ்பெல் உள்ளது. லேங்குவேஜ் டூல் ஒற்றுப்பிழையையும் சொற்களைச் சேர்த்தும் பிரித்தும் எழுதுவதால் ஏற்படும் தவறுகளையும் கண்டுபிடிக்கிறது. இவை இரண்டையும் ஒருங்கிணைத்தால் (integration) நமக்கு ஒரு முழுமையான தமிழ் பிழைத்திருத்தி கிடைக்கும் வாய்ப்பிருக்கிறது. [லேங்குவேஜ் டூல் ஜாவா மொழியில் எழுதி LGPL உரிமத்தின் கீழ் வெளியிடப்பட்டுள்ளது.](#) இதை ஒரு சவாலாக எடுத்துச்செய்ய தமிழ் சமூக நிரலாளர்கள் முன் வருவார்களா?

27. உணர்வு பகுப்பாய்வும் சமூக ஊடகங்களும்

உணர்வு பகுப்பாய்வு (sentiment analysis) அல்லது கருத்து சுரங்க வேலை (opinion mining) என்பது ஒரு பேச்சாளரின் அல்லது எழுத்தாளரின் மனோபாவத்தைத் தீர்மானிப்பது. ஒரு தலைப்பைப் பற்றியோ அல்லது ஒரு ஆவணத்தை ஒட்டுமொத்தமாகவோ ‘நேர்மறை (positive)’ அல்லது ‘எதிர்மறை (negative)’ என்று கணிக்கிறோம். இம்மாதிரி நேரெதிரான இரண்டு தன்மைகள் இருந்தால் அவற்றை முனைவு (polarity) என்று சொல்கிறோம். சில வேலைகளுக்கு மூன்றாவதாக ‘நடுநிலை (neutral)’ என்றும் கணிக்க வேண்டியிருக்கலாம். இது தவிர உயர்நிலை உணர்வு பகுப்பாய்வில் “கோபம்”, “சோகம்”, மற்றும் “மகிழ்ச்சி” போன்ற வகைப்பாடுகளும் செய்ய இயலும்.

இணையப் பக்கங்கள், செய்திகள், விவாதங்கள், விமர்சனங்கள், வலைப்பதிவுகள், மற்றும் சமூக ஊடகங்கள் உள்ளிட்ட ஆவணங்களின் பெரிய சேகரிப்புகளில் இம்மாதிரி உணர்வு பகுப்பாய்வு செய்ய வேண்டுமானால் கணினி உதவி தேவைப்படுகிறது. பொதுவாக சமூக ஊடகங்களில் யாவரும் தங்கள் கருத்துகளை வெளிப்படையாகப் பகிர்ந்து கொள்வதால், இம்மாதிரி கருத்து சுரங்க வேலை செய்யும் போது, சமூக ஊடக உள்ளடக்கம் மிகவும் பயனுள்ளதாக இருக்கிறது. மேலும் நிகழ் நேர கருத்துகள் தேவையெனில் ட்விட்டர் கீச்சுகளுக்கு நிகரில்லை.

ரசிகர்களின் கருத்துக் கணிப்பு

எடுத்துக்காட்டாக, இணையம் மற்றும் சமூக ஊடகங்கள் வருவதற்கு முந்தைய காலத்தில் ஒரு திரைப்படம் வெளியிடப்படுகிறது என்று வைத்துக் கொள்வோம். அந்தப் படத்தைப் பற்றி ரசிகர்களின் கருத்து என்ன என்று அதன் தயாரிப்பாளர்கள் அறிய விரும்பினால் முதல் காட்சி முடிந்து வெளியே வரும் ரசிகர்களிடம் நேரடியாகக் கருத்துக் கணிப்பு செய்வார்களாம். அதிகம் போனால் ஒருசில திரையரங்குகளில்தான் இப்படிச் செய்ய முடியும்.

இன்று ரசிகர்களோ தங்கள் விமர்சனங்களைப் படம் முடிந்தவுடன் (முடியும் முன்னரே?) கீச்சுத் தொடங்குகின்றனர். இன்று படங்களும் உலகம் முழுவதும் ஒரே நேரத்தில் வெளியிடப்படுகின்றன. கீச்சுகளை உணர்வு பகுப்பாய்வு செய்தால் கருத்துகளை உடன் அறிய முடியும். மக்களின் ரசனைக்குத் தகுந்தவாறு வெவ்வேறு வட்டாரங்களில் வெவ்வேறு வகையாக வினியோகமும் விளம்பரங்களும் செய்யலாம். திரைப்படங்களுக்கு மட்டுமல்ல, நுகர்பொருள்களுக்கும் ஆடை அலங்காரங்களுக்கும் மற்றும் பல ரசனை அடிக்கடி மாறும் வணிகப் பொருட்களுக்கும் சேவைகளுக்கும் இம்மாதிரி உணர்வு பகுப்பாய்வு தேவைப்படுகிறது.

தேர்தல் சமயங்களிலும் மற்ற நேரங்களிலும் அரசியல்வாதிகளும், அரசு சாரா நிறுவனங்களும் மற்றவர்களும் முறைசாரா கருத்துக் கணிப்புகளுக்கும் இம்மாதிரி செயல்முறைகளைப் பயன்படுத்தலாம்.

உணர்வு பகுப்பாய்வு



ஒரு தயாரிப்பு அல்லது சேவையைப் பற்றி மக்களின் கருத்துகளையும், உணர்ச்சிகளையும் இயல்மொழி ஆய்வு செய்து கண்டறிதல்

உணர்வு பகுப்பாய்வு

கருத்தைக் கணித்து அதற்கேற்ப தானியங்கியாக செயல்படுதல்

இரைச்சலை வடிகட்டுதல், உரையாடல்களைப் புரிந்துகொள்ளுதல், பொருத்தமான உள்ளடக்கத்தை அடையாளம் காணுதல் மற்றும் அதற்கேற்ப செயல்படுதல் போன்ற செயல்களைத் தானியங்கியாக வணிக நிறுவனங்கள் செய்ய முயற்சிக்கும் போது, அவர்களது கவனம் உணர்ச்சி பகுப்பாய்வு துறைக்கு திரும்பத் தொடங்கியுள்ளது.

மின்வணிக இணையதளங்களில் நீங்கள் ஒரு பொருளைத் தேடி எடுத்தீர்களானால், அந்தப் பொருளுடன் வாங்கக் கூடிய மற்ற பொருள்களையும் பரிந்துரை செய்கிறார்கள். இதைப் பரிந்துரை இயந்திரம் (Recommendation Engine) அல்லது அமைப்பு என்று சொல்கிறார்கள்.

பரிந்துரை அமைப்புக்கு, உணர்வு பகுப்பாய்வு ஒரு மதிப்புமிக்க நுட்பமாக நிரூபிக்கப்பட்டுள்ளது. பயனரின் விருப்பங்களை அறிந்து அதற்குத் தோதாக பரிந்துரை செய்ய இது உதவுகிறது.

நடைமுறை எடுத்துக்காட்டு: ட்விட்டர் கீச்சுகளை எடுத்து உணர்வு பகுப்பாய்வு

இந்தக் கானொலி ஆங்கில உரையைப் பற்றியது. இருப்பினும் தமிழில் விளக்கங்கள் எளிதாக இருப்பதால் இது நல்ல எடுத்துக்காட்டாக உதவும். ஒரு குறிப்பிட்ட தலைப்புக்கு கீச்சுகளை சேகரிக்க வேண்டுமானால் ட்விட்டரின் பயன்பாட்டு நிரலாக்க இடைமுகம் மூலம் எடுக்க முடியும். பைதான் நிரல் எழுதி எப்படி அதைச் செய்வது என்று இந்தக் கானொலியில் விளக்குகிறார்.

இதற்கு முன்தேவையாக apps.twitter.com சென்று நீங்கள் ஒரு செயலி உருவாக்கிக் கொள்ள வேண்டும். இந்த செயலியிலிருந்து நுகர்வோர் திறவி (API Key), நுகர்வோர் இரகசியம் (API Secret) அணுகல் அடையாளச் சின்னம் (Access Token), அணுகல் அடையாளச் சின்ன இரகசியம் (Access Token Secret) ஆகிய நான்கையும் படியெடுத்துக் கொள்ளவும். இவற்றை வைத்து பைதானிலுள்ள tweepy என்ற நிரலகத்தைப் பயன்படுத்தினால் நீங்கள் கொடுத்த சொல் பற்றிய கீச்சுகளை தேடித்தரும். இந்த கீச்சுகளை ஒன்றொன்றாக TextBlob என்ற நிரலகத்தைப் பயன்படுத்தி உணர்ச்சி பகுப்பாய்வு செய்ய முடியும்.

ஆங்கிலத்தில் இயல்மொழி ஆய்வுக்கான வளங்கள் மிகுதியாக உள்ளன

இது எப்படி ஒரே வரியில் பகுப்பாய்வு செய்ய முடிந்தது, மாய வேலை போலிருக்கிறதே என்கிறீர்களா? இதனால்தான் ஆங்கிலத்தில் வளங்கள் அதிகமாக உள்ளன, ஆனால் தமிழில் மிகக்குறைவு என்று சொல்கிறோம். இந்த வேலையைச் செய்ய TextBlob ஒரு அகராதியுடன் வருகிறது.

இந்த அகராதியில் ஒவ்வொரு சொல்லும் கீழ்க்கண்ட மதிப்பெண்களைக் கொண்டுள்ளது.

- முனைவு: எதிர்மறையா அல்லது நேர்மறையா
- அகநிலைப்பாடு: தற்சார்புடைய (subjective) அல்லது தற்சார்பற்ற (objective)
- தீவிரம் (intensity): அடுத்த சொல்லை மாற்றியமைக்கிறது (எடுத்துக்காட்டாக, “மிக” ஒரு மாற்றியமைக்கும் சொல்)

ஒரு தனிப்பட்ட குறிப்பு. பகுப்பாய்வு செய்யும்போது TextBlob ஓரெழுத்துச் சொற்களைப் பொருட்படுத்தாது. மற்ற சில இயல்மொழி ஆய்வுகளில் ஈரெழுத்துச் சொற்களையும் விட்டு விடுகிறார்கள். ஏனெனில் ஆங்கிலத்தில் இவை வாக்கியத்தின் பொருளுக்கு அவசியமில்லாத stop words என்று சொல்கிறார்கள். ஆனால் தமிழிலோ ஆ, ஈ, தீ, கை, மா, வா, போ, வை போன்ற ஓரெழுத்துச் சொற்கள் மிக அவசியமானவை. எனினும் ஆங்கிலத்துடன் ஒப்பிடுகையில் ஆ, ஈ மட்டுமே ஓரெழுத்துச் சொற்கள். ஒரு மெய்யெழுத்தும் ஒரு உயிரெழுத்தும் சேர்ந்துள்ளதால் மற்ற உயிர்மெய் எழுத்துகள் ஈரெழுத்துச் சொற்களே.

தேசிய அறிவியல் அறக்கட்டளை (NSF) மானியமும், ஆல்ஃபிரெட் பி. ஸ்லோன் ஆய்வு உதவித் தொகையும் பெற்றுகார்னெல் பல்கலையில் சில ஆய்வாளர்கள் உணர்வு பகுப்பாய்வு சோதனைகளில் பயன்படுத்த தரவுகளைத் தயாரித்தனர். இந்த தரவுத்தளத்தில் இணைய திரைப்பட தரவுத்தளம் (IMDB), rec.arts.movies.reviews செய்திக்குழு, அழகிய தக்காளிகள் (Rotten Tomatoes) திரைப்படங்கள் மற்றும்

தொலைக்காட்சிகளை மதிப்பாய்வு மற்றும் திரட்டு செய்யும் அமெரிக்க இணையதளம் ஆகியவற்றிலிருந்து எடுக்கப்பட்ட விமர்சனங்கள் உள்ளன.

- 1000 நேர்மறை மற்றும் 1000 எதிர்மறை சீர் செய்யப்பட்ட திரைப்பட விமர்சனங்கள்.
- 5331 நேர்மறை மற்றும் 5331 எதிர்மறை சீர் செய்யப்பட்ட வாக்கியங்கள் மற்றும் துண்டுகள்.
- 5000 தற்சார்புடைய (subjective) மற்றும் 5000 தற்சார்பற்ற (objective) சீர் செய்யப்பட்ட வாக்கியங்கள்.

சமூக ஊடகத் தரவுகளை இயல்மொழி ஆய்வு செய்வதில் எழும் சவால்கள்

சமூக இயங்குதளங்களின் கட்டுப்பாடற்ற சூழல்கள் பயனர்கள் பேச்சுவழக்கு, தனிப்பட்ட சொற்கள் மற்றும் அஃகுப்பெயர்களைப் (acronyms) பயன்படுத்துவதை ஊக்குவிக்கின்றன. சில எடுத்துக்காட்டுகள் ஜுஜுபி, மரண மொக்கை, ROFL ஆகியவை. இவற்றைப் புரிந்து கொள்வதற்காக தமிழ் அகராதியில் சேர்ப்பதா, ROFL போன்ற சுருக்கங்களை முன்செயலாக்கத்தில் விரித்து எழுதி மொழிபெயர்ப்பு செய்வதா போன்ற பிரச்சினைகள் சமூக ஊடகத் தரவுகளில் எழும். பெரும்பாலும் இவற்றை உணர்வு பகுப்பாய்வு திட்டங்களால் புரிந்து கொள்ள முடிவதில்லை. ஆகவே சூழல் மற்றும் உணர்வை விளக்குவது கடினமாக உள்ளது.

இருப்பினும், நிறுவனங்கள் தங்கள் வாடிக்கையாளர்களுடன் நேரடியாகத் தொடர்பு கொள்ளவும், அவர்களின் விருப்பங்களைப் புரிந்துகொள்ளவும் சமூக ஊடகங்கள் மிகச் சிறந்த வழிமுறையாகவே உள்ளன. திறமையான முறையில் கையாண்டால், அது நீண்ட கால வணிக வளர்ச்சிக்கு நுண்ணறிவு உருவாக்க பயன்படும் தரவுகளை சேகரிக்க நிறுவனங்களுக்கு ஒரு சக்திவாய்ந்த கருவியாகும்.

28. மொழித் தொழில்நுட்பத்தில் வளங்கள் மிகுந்த மொழியாகத் தமிழை உயர்த்துவோம்

இயல் மொழியியலில் அண்மைய தொழில்நுட்பக் கலை பற்றிய ஆய்வு, தொகுதி 13-14 இலிருந்து கீழ்க்கண்ட மேற்கோள் எடுக்கப்பட்டது.

“மொழித் தொழில்நுட்பத்தை பயன்படுத்துவதன் மூலம் பெறக்கூடிய நன்மைகள் இவை. கணினிகளின் பயன்மை (usability) அதிகரிக்கிறது. மேலும் கணினி பயன்பாட்டில் பாமர மக்கள் தன்மேம்பாடு பெறவும் (empowering) வழிவகுக்கிறது.”

தமிழின் கடந்த முதன்மைத்துவத்தை மீண்டும் பெற முயல்வோம்

1805 இல் ராஜேட் ஆங்கிலத்தில் முதல் தெஸாரஸ் (Roget's Thesaurus) உருவாக்கினார். இதன் மூலப் பதிப்பில் 15,000 சொற்கள் இருந்தன. அச்சிடல் ஐரோப்பாவில் 15 ஆம் நூற்றாண்டிலேயே பரவலாக வந்துவிட்டது. திவாகர நிகண்டு என்னும் நூல் 8 ஆம் நூற்றாண்டில் திவாகர முனிவரால் இயற்றப்பட்டது. பிங்கள நிகண்டு 10 ஆம் நூற்றாண்டில் இயற்றப்பட்டது. இதில் 14700 சொற்பொருள் விளக்கங்கள் உள்ளன. நிகண்டு என்பது ஒத்த கருத்துடைய சொற்களை ஒரிடத்தில் சேர்த்துக் கொடுக்கும் தெஸாரஸ் போன்ற நூல் ஆகும். அக்காலத்தில் ஓலைச்சுவடியும் பரவலாக வரவில்லையென்று தெரிகிறது. இந்நூற்களை அனேகமாக மனப்பாடமாகவே பல தலைமுறைகளுக்குப் பேணிக்காத்தனர். கணினி மற்றும் தகவல் யுகத்தில் நம் தாய்மொழியைத் தழைக்கச் செய்யத் தேவையான வேலைகளை நாம் செய்யாவிட்டால் என்ன சாக்கு சொல்லமுடியும்? நமக்கு என்ன வசதி இல்லை? நம் முன்னோர்களிடமிருந்த மொழிப்பற்றையும் சமூக மனநிலையையும் அடைய முயல்வோம்.

இயல்மொழி ஆய்வில் தமிழுக்கு பலவீனமான ஆதரவுதான் உள்ளது

உலகத்தில் சுமார் 7000 மொழிகள் உள்ளன. இவற்றை இயல்மொழி ஆய்வின் ஆதரவு மற்றும் வளங்கள்படி ஆறு படிக்களாகப் பிரித்தார்கள். கருவிகளும் வளங்களும் மிகுந்து முதல் படியில் உன்னதமான ஆதரவுடன் ஆங்கிலம் தனியிடத்தில் உள்ளது. இரண்டாம் இடத்தில் நல்ல ஆதரவுடன் ஜெர்மன், பிரெஞ்சு, ஸ்பானிஷ், இத்தாலிய, போர்த்துகீய மொழிகள் போன்ற 10 உள்ளன. மூன்றாவது இடத்தில் மிதமான ஆதரவுடன் இந்தி, வியட்நாமிய, தாய்லாந்து போன்ற 70 மொழிகள் உள்ளன. தமிழ் நான்காவது இடத்தில் பலவீனமான ஆதரவுடன் 300 மொழிகளில் ஒன்றாகத்தான் இன்று இருக்கிறது. 330 ஆயிரம் பேருக்கு மட்டுமே தாய்மொழியான ஐஸ்லாண்டு மொழிக்கு நம்மைவிட நல்ல கருவிகளும் வளங்களும் உள்ளன!

மொழிக்கான அடிப்படை வளங்கள் கருவிப் பெட்டி

ஐரோப்பிய மொழிகளுக்கு, மிகச்சிறு மக்கள்தொகை மட்டுமே பேசும் மொழிகளுக்குக் கூட, தொழில்நுட்பமும் வளங்களும் இன்று பெருகியிருக்கக் காரணம் தெரியவேண்டுமென்றால், அவர்கள் பல பத்தாண்டுகளாக திட்டமிட்டு எடுக்கும் முயற்சிகளைப் பார்க்க வேண்டும். இவற்றில் முக்கியமான ஒன்று மொழிக்கான அடிப்படை வளங்கள் கருவிப் பெட்டி (Basic Language Resource Kit – BLARK) தயார் செய்தல்.

ஐரோப்பாவில் 11 அதிகாரப்பூர்வமான மொழிகளும் 200 க்கும் மேற்பட்ட மற்ற மொழிகளும் உள்ளன. அடிப்படைவளங்கள் கருவிப் பெட்டி என்ற கட்டமைப்பு ஸ்டீவன் கிரௌவர் (Steven Krauwer) என்பவரால் ஐரோப்பிய மொழிகளுக்காக முன்வைக்கப்பட்டது. இதன் குறிக்கோள் ஒரு மொழியைக் கணினிக்கு கிடைக்க வைப்பது (making a language computationally available) என்று சொல்கிறார்கள். அதாவது இயல்மொழியை ஆய்வு செய்ய கருவிகள், தரவுகள், செயல்முறைகள் போன்றவை தேவையான அளவு இருந்தால்தான் கணினி உதவி ஆய்வுகள் செய்ய இயலும்.

கல்வி மற்றும் வணிகத்துறை ஆராய்ச்சியாளர்கள், மற்றும் கல்வியாளர்கள் இந்த கருவிப் பெட்டியின் இலக்கு பயனர்கள். இது மாணவர்களுக்குப் பயிற்சியளிக்கவும், ஆராய்ச்சி சோதனைகளுக்கும், செயலி

வெள்ளோட்டங்களுக்கும் பயன்படுத்தப்படும். இந்தக் கருவிப் பெட்டியை வணிக நிறுவனங்கள் பயன்படுத்தலாம், ஆனால் பொதுவாக வணிக வேலைகளுக்கு மாற்றமின்றி அப்படியே பயன்படுத்துவது அனேகமாக சாத்தியமில்லை.

எனவே இந்தக் கருவிப் பெட்டி திறந்த மூலமாகவும் இதற்கான வளங்கள் திறந்த அனுமதியிலும் கிடைப்பது இன்றியமையாத முக்கியத்துவம் வாய்ந்தது. அப்படிச் கிடைத்தால்தான் தொழில்துறை உருவாக்குநர்கள் தயக்கமின்றித் தங்கள் தேவைகளுக்கு ஏற்ப இவற்றை மாற்றிக்கொண்டு பயன்படுத்த முடியும்.

நடைமுறைக்கேற்ற, அடையக்கூடிய குறிக்கோள்

ஆக இயல்மொழி ஆய்வுகளில் தமிழ் இன்று வளங்கள் வறண்ட மொழியாக இருக்கிறது. இதை ஆங்கிலத்துக்குச் சமமாக ஆக்குவது அடையக்கூடிய குறிக்கோள் அல்ல. வளங்கள் மிகுந்த மொழியாக ஆக்கக்கூடும். ஆனால் இப்போது அதை நினைத்துப் பார்க்கக் கூடிய இடத்தில் நாம் இல்லை. இடைக்காலத்தில், அதாவது மூன்று முதல் ஐந்து ஆண்டுகளுக்குள், மிதமான வளங்கள் கொண்ட மொழியாக ஆக்குவதுதான் நடைமுறைக்கேற்ற, அடையக்கூடிய ஆனால் சவாலான குறிக்கோள். இதற்கு முதல் படியாக நமக்கு நல்ல தரமான மொழித் தொகுப்புகள் படைப்பாக்கப் பொதுமங்கள் அனுமதியில் தேவை. அடுத்து திறந்த மூலமாக ஒரு இயல்மொழி ஆய்வு கருவித்தொடரி தேவை. மேலும், பிழைகளற்ற உரைத்தொகுப்புகள் தயாரிக்க பிழைத்திருத்தியும் திறந்த மூலமாக அவசியம். இவற்றைக் குறுகிய காலத்தில், ஓராண்டுக்குள், தயாரிக்க வேண்டும்.

இயல்மொழி ஆய்வுக் கட்டுரைகள் தமிழில்

முரண்நகையாக தமிழ் மொழி பற்றிய இயல்மொழி ஆய்வுக் கட்டுரைகள் ஆங்கிலத்தில் மட்டுமே வெளியிடப்படுகின்றன! தமிழில் மதிப்புமிக்க ஆய்விதழ்கள் (reputed research journals) இல்லை, மேற்கோள் (citation) காட்ட ஆங்கிலத்தில் வெளியிட்டால்தான் வசதியாக இருக்கும் என்பது எனக்குப் புரிகிறது. ஆனால் தமிழிலும் வெளியிடக்கூடாதா என்பதுதான் என்னுடைய கேள்வி. தமிழில் வெளியிட நீங்கள் முன்வந்தால் கணியம் இதழ் வெளியிடத் தயாராக உள்ளது. மொழிபெயர்ப்பையும் நாங்களே செய்துகொள்கிறோம். ஆங்கிலக் கட்டுரையை மட்டும் அனுப்புங்கள்.

பண்பாட்டுக்குத் தமிழ்; பணிவாய்ப்புக்கு ஆங்கிலம்

ஒரு உலகளாவிய தகவல்தொடர்பு சூழலில், வணிக மற்றும் பணிவாய்ப்புகளை விரிவாக்க நாம் பரவலாக யாவருடனும் தொடர்பு கொள்ள வழிவகை தேவை. அதே வேளையில் நமது தாய்மொழியையும், அத்துடன் சேர்ந்த நமது கலாச்சார அடையாளத்தையும் பாதுகாத்துக் கொள்வதற்கான வழிகளையும் நாம் கண்டறிய வேண்டும். “பண்பாட்டுக்குத் தமிழ்; வணிக மற்றும் பணிவாய்ப்புக்கு ஆங்கிலம்” இதுதான் தமிழர்களின் இன்றைய மொழி பற்றிய அணுகுமுறையாக இருக்க முடியும். வணிக மற்றும் பணிவாய்ப்புக்காக ஆங்கிலக் கல்விக்கு முன்னுரிமை கொடுக்கும் நாம், நம் பிள்ளைகளுக்குத் தாய் மொழியான தமிழில் ஆர்வம் வளர்க்கவில்லையென்றால் அவர்கள் கலாச்சாரமும் பண்பாடுமற்ற இயந்திர மனிதர்களாகவே வளர்வார்கள்.



நீங்களும் பங்களிக்க முடியும்

ஊர் கூடித்தான் தேரிழுக்க முடியும்

இந்த திட்டம் நம் தாய் மொழியைப் பிழைக்க வைப்பது பற்றியது மட்டுமல்ல, செழிக்க வைப்பது பற்றியதும்தான். நாம் யாவரும் ஒரு கை கொடுத்தால் சாதிக்க முடியாதது எதுவுமில்லை. உங்களுக்கு நிரல் எழுதத் தெரியாதா? சோதனை செய்வது, ஆவணங்கள் எழுதுவது, உரைகளைத் தொகுப்பது, வலைப்பதிவு எழுதுவது, தமிழாக்கம் செய்வது போன்ற பல வேலைகள் உள்ளன. நீங்களும் தமிழை செழிக்க வைக்கப் பங்களிக்க முடியும்.

முற்றும்